Mickelsson lowering operators for the symplectic group

Adam M. Bincer
Physics Department, University of Wisconsin—Madison, Madison, Wisconsin 53706

(Received 21 September 1981; accepted for publication 20 November 1981)

Elementary lowering operators for the symplectic group, for which a graphical algorithm was given by Mickelsson, are obtained in the form of tensor operators. The resultant simple analytic expressions are analogous to the corresponding ones found previously for the unitary and orthogonal groups.

PACS numbers: 02.20.Qs

1. INTRODUCTION

Shift operators have been introduced for the purpose of constructing a basis in the representation space of a group. By definition, a basis vector of any weight can be obtained from a basis vector of any other weight by the application of an appropriate shift operator. In particular, if the representation space contains a maximal vector the entire representation space may be obtained by applying appropriate lowering operators to the maximal vector.

These ideas were first formulated explicitly in application to the unitary groups by Nagel and Moshinsky,[1] who gave analytic expressions for the shift operators. Until recently attempts to extend these results to other groups have proceeded in two ways. On the one hand, analytic results were obtained for the SO(n) and SO(n, 1) groups by Wong,[2] for the U(n, 1) groups by Patera,[3] and for the discrete series of the U(p, q) groups by Patera, Winternitz, and Sharp.[4] On the other hand, less explicit results based on a graphical algorithm were obtained for the SO(n) groups by Pang and Hecht,[5] and for the Sp($2n$) groups by Mickelsson.[6]

Recently I have proposed another approach, which makes use of the concept of tensor operators due to Louck and Biedenharn.[7] In this approach the nonuniqueness inherent in the definition of the shift operators is exploited to express them as tensor operators. By this means much simpler analytic expressions were obtained for the U(n) and SO(n) groups,[8] as well as for the discrete series of the SO(p, q) groups.[9] In the present work the tensor operator technique is used to obtain simple analytic expressions for the elementary lowering operators of the Sp($2n$) groups, thus closing the gap in the list of the classical groups.

2. TENSOR OPERATORS

Following Mickelsson[6] I denote the generators of Sp($2n$) by $E_{ij} = E_{ji}$ with indices ranging from $-n$ to n, zero excluded. The Sp($2t$) subgroup, $1 \leqslant t < n$, is obtained by restricting the indices to range from $-t$ to t, zero excluded.

The commutation relations of the generators are

$$[E_{ij}, E_{kl}] = g_{ik}E_{jl} + g_{jk}E_{il} + g_{jl}E_{ki} + g_{il}E_{kj}, \tag{2.1}$$

with the symplectic metric given by

$$g_{ij} = 0 \quad \text{if } i \neq -j \quad \text{and } g_{-ii} = -g_{i-i} = i/|i|. \tag{2.2}$$

In a unitary representation the generators have the Hermiticity property

$$E_{ij}^{\dagger} = -g_{ik}g_{jl}E_{kl}. \tag{2.3}$$

I define an Sp($2t$) tensor operator T, with $(2t)^2$ components T_{ij}, $1 \leqslant |i|, |j| \leqslant t$, by the transformation properties

$$[E_{ij}, T_{kl}] = g_{ik}T_{jl} + g_{jk}T_{il} + g_{jl}T_{ki} + g_{il}T_{kj}. \tag{2.4}$$

It is obvious that E, whose components are the generators, is a tensor operator. Another example is the symplectic metric g (an invariant tensor operator). It is easily shown that if T and U are two Sp($2t$) tensor operators then so is TU, where

$$(TU)_{ij} \equiv -T_{ik}U_{ij}g_{kl} \tag{2.5}$$

with all indices in the range appropriate to Sp($2t$).

3. MAXIMAL AND SEMIMAXIMAL VECTORS

The Cartan subalgebra is generated by the n $E_{i-i} \equiv H_i$ (no summation), $1 \leqslant i \leqslant n$. A vector $|\mathbf{w}\rangle$ in the representation space may be taken to be a simultaneous eigenvector of the H_i:

$$H_i|\mathbf{w}\rangle = w_i|\mathbf{w}\rangle, \quad 1 \leqslant i \leqslant n, \tag{3.1}$$

$$\mathbf{w} \equiv (w_1, w_2, \ldots, w_n), \tag{3.2}$$

where $\mathbf{w}$ is called the weight. Two weights are called equal if all the components of the weights are equal, otherwise the weight $\mathbf{w}'$ is called higher than $\mathbf{w}$ if

$$w_i' = w_i \quad \text{for } 1 \leqslant i < j \quad \text{and } w_j' > w_j. \tag{3.3}$$

The T_{ij} belong to the set $\alpha^+(\alpha^-)$ of *raising* (*lowering*) tensor operators if in the equation

$$T_{ij}|\mathbf{w}\rangle = |\mathbf{w}'\rangle \tag{3.4}$$

the weight $\mathbf{w}'$ is higher (lower) than $\mathbf{w}$. It follows that

$$\alpha^+ = \{T_{ij} | i, j > 0\} \cup \{T_{i-j} | j > i > 0\}, \tag{3.5}$$

$$\alpha^- = \{T_{ij} | i, j < 0\} \cup \{T_{i-j} | i > j > 0\}. \tag{3.6}$$

Since all unitary irreducible representations (unireps) of Sp($2n$) are finite-dimensional they necessarily contain a basis vector of highest weight, the so-called *maximal* vector $|\mathbf{m}^{2n}\rangle$:

$$E_{ij}|\mathbf{m}^{2n}\rangle = 0, \quad E_{ij} \in \alpha^+, \quad 1 \leqslant |i|, |j| \leqslant n, \tag{3.7}$$

$$H_i|\mathbf{m}^{2n}\rangle = m_i^{2n}|\mathbf{m}^{2n}\rangle, \quad 1 \leqslant i \leqslant n, \tag{3.8}$$

and this highest weight characterizes the unirep. Since a unirep of Sp($2n$) is necessarily a representation (possibly reducible) of the Sp($2n-2$) subgroup, the representation space of Sp($2n$) characterized by $\mathbf{m}^{2n}$ must also contain so-called *semimaximal* vectors |s.m.⟩, which are maximal vectors of

 0022-2488/82/030347-03$02.50 © 1982 American Institute of Physics

Sp(2n − 2):

$$E_{ij}|\text{s.m.}\rangle = 0, \quad E_{ij}\in\alpha^+, \quad 1\leqslant|i|,|j|\leqslant n-1, \tag{3.9}$$

$$H_i|\text{s.m.}\rangle = m_i^{2n-2}|\text{s.m.}\rangle, \quad 1\leqslant i\leqslant n-1, \tag{3.10}$$

$$H_n|\text{s.m.}\rangle = w_n|\text{s.m.}\rangle. \tag{3.11}$$

4. THE MICKELSSON LOWERING OPERATORS

A basis in the Sp(2n) representation space may be constructed by determining the lowering operators, whose application to the maximal vector yields all independent semimaximal vectors, and then iterating this procedure down the subgroups chain $\text{Sp}(2n)\supset\text{Sp}(2n-2)\supset\cdots\supset\text{Sp}(2)$. Such lowering operators were given by Mickelsson[6] in the form of a polynomial in E_{-n-n} and $L_j^{\pm n}$, $1\leqslant j\leqslant n-1$. The $L_j^{\pm n}$ are the elementary lowering operators defined by[6]

$$[H_n, L_j^{\pm n}] = \pm L_j^{\pm n}, \tag{4.1}$$

$$[H_k, L_j^{\pm n}] = -\delta_{kj}L_j^{\pm n}, \quad 1\leqslant k\leqslant n-1, \tag{4.2}$$

$$[E_{ik}, L_j^{\pm n}]|\text{s.m.}\rangle = 0, \quad E_{ik}\in\alpha^+, \quad 1\leqslant|i|,|k|\leqslant n-1. \tag{4.3}$$

These equations state that the action of $L_j^{\pm n}$ on a semimaximal vector in Sp(2n) representation space is to produce another such semimaximal vector, whose jth weight is lower by 1, and whose nth weight is raised (for L_j^{+n}) or lowered (for L_j^{-n}) by 1.

To solve these equations Mickelsson[6] devised an algorithm using graphs. I proceed instead as follows. It is seen from the definition of tensor operators, Eq. (2.4), that Eqs. (4.1) and (4.2) are satisfied by

$$L_j^{+n} = \{E\,T(j)\}_{+n-j}, \quad 1\leqslant j\leqslant n-1 \tag{4.4}$$

where $T(j)$ is an *arbitrary* Sp(2n − 2) tensor operator. Moreover, Eq. (4.3) is also satisfied if the operator $T(j)$ obeys

$$T(j)_{lk}|\text{s.m.}\rangle = 0 \quad \text{for } 1\leqslant k\leqslant n-1 \quad \text{and } j<-k\leqslant n-1. \tag{4.5}$$

It is shown in the Appendix that a solution to Eq. (4.5) is given by

$$T(j) = \prod_{s=1}^{n-1}(E+d_s g)\prod_{p=j+1}^{n-1}(E+c_p g), \tag{4.6}$$

where

$$d_s \equiv s-1-m_s^{2n-2}, \tag{4.7}$$

$$c_p \equiv 2n-p-1+m_p^{2n-2}, \tag{4.8}$$

and where I use the convention

$$\prod_{p=n}^{n-1}(E+c_p g)\equiv g. \tag{4.9}$$

The reader is reminded that all products are defined by Eq. (2.5) with repeated indices summed over the Sp(2n − 2) range.

APPENDIX

If follows from Eqs. (4.6) and (2.5) that

$$T(j-1)_{lk} = \{T(j)(E-c_j g)\}_{lk} = c_j T(j)_{lk} + \sum_{p=1}^{n-1}\{T(j)_{lp}E_{-pk} - T(j)_{l-p}E_{pk}\}. \tag{A1}$$

Therefore, for $j-1<-k\leqslant n-1$,

$$\begin{aligned} T(j-1)_{lk}|\text{s.m.}\rangle &= (c_j - m_{-k}^{2n-2})T(j)_{lk}|\text{s.m.}\rangle + \sum_{p=1}^{n-1}\{E_{-pk}T(j)_{lp} + [T(j)_{lp}, E_{-pk}]\}|\text{s.m.}\rangle \\ &\quad - \sum_{p=-k+1}^{n-1}\{E_{pk}T(j)_{l-p} + [T(j)_{l-p}, E_{pk}]\}|\text{s.m.}\rangle \\ &= (c_j - m_{-k}^{2n-2} - 2n - k + 1)T(j)_{lk}|\text{s.m.}\rangle + \sum_{p=1}^{n-1}\{E_{-pk}T(j)_{lp} - g_{-pl}T(j)_{kp} - g_{kl}T(j)_{-pp}\}|\text{s.m.}\rangle \\ &\quad - \sum_{p=-k+1}^{n-1}\{E_{pk}T(j)_{l-p} - g_{pl}T(j)_{k-p} - g_{kl}T(j)_{p-p}\}|\text{s.m.}\rangle, \end{aligned} \tag{A2}$$

where I have used the fact that Sp(2n − 2) raising tensor operators annihilate the semimaximal vector.

Assuming Eq. (4.5) to be true, I deduce that the right-hand side of Eq. (A2) vanishes for $j<-k\leqslant n-1$. Moreover, the right-hand side of Eq. (A2) also vanishes for $j=-k$ by definition of c_j, Eq. (4.8). By similar, except simpler, manipulations I deduce from Eq. (A1) that $T(j-1)_{lk}|\text{s.m.}\rangle$ also vanishes for $1\leqslant k\leqslant n-1$. Thus I have shown that the validity of Eq. (4.5) implies the validity of the same equation with j replaced by $j-1$. It then follows by iteration that Eq. (4.5) is true for $1\leqslant j\leqslant n-2$ provided it is true for $j=n-1$, i.e., provided

$$T(n-1)_{lk}|\text{s.m.}\rangle = 0 \quad \text{for } 1\leqslant k\leqslant n-1, \tag{A3}$$

where I leave off the restriction $j<-k\leqslant n-1$ because it is empty for $j=n-1$.

To prove Eq. (A3) let

$$W(p)\equiv\prod_{s=1}^{p-1}(E+d_s g), \quad 2\leqslant p\leqslant n \tag{A4}$$

so that

$$T(n-1) = W(n). \tag{A5}$$

Using the same technique as above I can show that the equation

$$W(p)_{lk}|\text{s.m.}\rangle = 0 \quad \text{for } 1\leqslant k\leqslant p-1 \tag{A6}$$

implies the validity of the same equation with p replaced by

$p+1$. Hence by iteration Eq. (A6) is true for $2<p\leqslant n$ provided it is true for $p=2$. The proof is completed by observing that

$$W(2)_{l1}\,|\text{s.m.}\rangle = (E+d_1 g)_{l1}\,|\text{s.m.}\rangle = \delta_{l-1}(m_1^{2n-2}+d_1)|\text{s.m.}\rangle = 0. \quad \text{(A7)}$$

ACKNOWLEDGMENTS

This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the Department of Energy under contract DE-AC02-76ER00881, COO-881-142.

[1] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); Rev. Mex. Fis. **14**, 29 (1965).
[2] M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967); **15**, 25 (1974).
[3] J. Patera, J. Math. Phys. **14**, 279 (1973).
[4] J. Patera, P. Winternitz, and R. T. Sharp, Rev. Mex. Fis. **23**, 81 (1974).
[5] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[6] J. Mickelsson, Rep. Math. Phys. **3**, 193 (1972).
[7] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[8] A. M. Bincer, J. Math. Phys. **18**, 1870 (1977); **19**, 1173 (1978); **19**, 1179 (1978).
[9] T. Murphy, J. Math. Phys. **21**, 2317 (1980).

Boson–fermion representations of Lie superalgebras: The example of osp(1,2)

Jiří Blank, Miloslav Havlíček
Nuclear Centre, Faculty of Mathematics and Physics, Charles University, Prague, Czechoslovakia

Pavel Exner[a)]
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, USSR

Wolfgang Lassner
Sektion Physik, Karl-Marx-Universität, Leipzig, German Democratic Republic

(Received 25 February 1980; accepted for publication 27 February 1981)

A method for constructing infinite-dimensional representations of Lie superalgebras employing boson representations of their Lie subalgebras is outlined. As an example the osp(1,2) superalgebra is considered; explicit formulae for its generators in terms of one pair of boson operators, at most one pair of fermion ones, and at most one parameter are obtained, the Casimir operator being represented by a multiple of unity. The restriction of these representations to the real form of osp(1,2) is skew-symmetric in the even part and can be regarded as a natural generalization of skew-symmetric representations of real Lie algebras. Some other aspects of the presented construction are discussed.

PACS numbers: 02.20.Qs, 11.30.Pb, 02.20.Sv, 02.20.Tw

I. INTRODUCTION

1.1. The physical concept of supersymmetry has attracted attention to the mathematical structure of Lie superalgebras and their representations. The (finite-dimensional) Lie superalgebras can be classified in a manner analogous to the Lie algerbra theory. A complete classification is known for the complex simple Lie superalgebras[1–6] as well as for real ones[1]; it allows one to describe also all semisimple Lie superalgebras, however, in a more complicated way than in the usual Lie theory. Various results are known concerning finite-dimensional irreducible representations of simple Lie superalgebras[1,7–10]; the situation is much less satisfactory for infinite-dimensional representations. In this paper we propose a method for constructing such representations and illustrate it on the simplest nontrivial example of the osp(1,2) superalgebra [called also B(0,1) or (sp(2),2) [5] and osp (2,1) [7]].

1.2. There exists a wide family of the so-called boson representations for real classical simple Lie algebras. They are obtained with the help of canonical realizations, i.e., realizations in terms of several pairs of boson creation and annihilation operators satisfying the canonical commutation relations (cf. Ref. 11 and references quoted therein). All of these representations are Schurean, i.e., every Casimir operator is represented by a multiple of unity. They are, moreover, skew-symmetric. Our aim is to use some of these boson representations for constructing representations of real Lie superalgebras which have the even elements skew-symmetric (this property will be hereafter referred to as ESS) and are Schurean. There are two reasons why we restrict ourselves to Schurean representations: (i) it is relatively simple to construct them by means of our method, whereas the problem of irreducibility is much more complicated as we are dealing with unbounded operators, (ii) non-Schurean representations are reducible.

1.3. The construction for a Lie superalgebra $L = L_0 \oplus L_1$ proceeds as follows. In a suitably chosen skew-symmetric boson representation of the Lie algebra L_0, which depends on at least one numerical parameter, we replace all the parameters by operators on some "fermion" Hilbert space $\mathscr{H}_F$. Then we look for generators of the odd part L_1 in the form of linear combinations of operators $T_B \otimes T_F$; $T_B \in \mathrm{End}\,\mathscr{H}_B$ depends on a certain number of pairs of operators on the representation space $\mathscr{H}_B$ of the boson representation satisfying the canonical commutation relations (CCR), and analogously $T_F \in \mathrm{End}\,\mathscr{H}_F$ is expressed by means of one or more pairs of operators on $\mathscr{H}_F$ which obey the canonical anticommutation relations (CAR). The trivial case $\dim\mathscr{H}_F = 1$ is not excluded; it yields the special type of "pure boson representations" of L where no fermion operators occur.

1.4. It should be stressed that not every skew-symmetric boson representation of L_0 is suitable for applying this approach. The choice is restricted especially by the ESS requirement. An example of a boson representation of sp(2, $\mathbb{R}$) which is not "suitable" just for this reason is given in Sec. 3. On the other hand, our results in the next section concerning the osp(1,2) superalgebra as well as our preliminary results for osp(1,4) suggest that there exists at least one suitable boson representation of L_0 which, moreover, gives rise to a family of representations of L depending on n parameters, n being the number of independent Casimir operators of L.

II. SCHUREAN REPRESENTATIONS OF THE osp(1,2) SUPERALGEBRA

In this section we shall apply the construction sketched in the introduction to the particular case of osp(1,2).

2.1. In the basis of osp(1,2) considered in Ref. 7 the even generators Q_+, Q_-, Q_3 and the odd ones V_+, V_- fulfill the relations

[a)] On leave from Faculty of Mathematics and Physics of the Charles University, Prague, Czechoslovakia.

 0022-2488/82/030350-04$02.50 © 1982 American Institute of Physics

$$[Q_3,Q_\pm] = \pm Q_\pm, \quad [Q_+,Q_-] = 2Q_3, \tag{1a}$$

$$[Q_3,V_\pm] = \pm \tfrac{1}{2}V_\pm, \tag{1b}$$

$$[Q_+,V_-] = V_+, \quad [V_\pm,V_\pm]_+ = \pm \tfrac{1}{2}Q_\pm,$$

$$[V_+,V_-]_+ = -\tfrac{1}{2}Q_3, \tag{1c}$$

$$[Q_\pm,V_\pm] = 0, \quad [Q_-,V_+] = V_-. \tag{1d}$$

The Casimir operator is expressed as

$$K_2 = Q_3^2 + \tfrac{1}{2}(Q_+Q_- + Q_-Q_+) + V_+V_- - V_-V_+. \tag{2}$$

2.2. The even part of osp(1,2) is the Lie algebra $sp(2,\mathbb{R}) \simeq sl(2,\mathbb{R})$. A suitable boson representation of this algebra can be constructed using one pair of boson operators b,b^+ which obey the CCR: $[b,b^+] = I_B$. It is convenient to pass to the linear combinations $p = 2^{-1/2}(b - b^+)$, $q = 2^{-1/2}(b + b^+)$ fulfilling

$$[p,q] = I_B. \tag{3}$$

We choose the operators p,q in such a way that p is skew-symmetric and q symmetric. This choice is consistent with the standard relation $b^+ \subset b^*$, where b^* is the adjoint of b. One can easily verify that the operators

$$Q^B_+ = iq^2, \quad Q^B_3 = \tfrac{1}{2}(qp + \tfrac{1}{2}I_B), \quad Q^B_- = \tfrac{1}{4}ip^2 + i\gamma q^{-2} \tag{4}$$

satisfy for each $\gamma \in \mathbb{C}$ the relations (1a) and generate a Schurean representation of $sp(2,\mathbb{R})$. Moreover, if $\gamma \in \mathbb{R}$, then this representation is skew-symmetric.

2.3. The relations (1a) are to be understood in the sense commonly accepted in the representation theory (cf. Ref. 12): there is a common dense invariant domain $D \subset \mathscr{H}_B$, i.e., such that all the operators (4) map D into D. One has, of course, to specify a representation of the CCR (3) appropriate for this case. In the following we put $\mathscr{H}_B = L^2(\mathbb{R})$ and use the Schrödinger representation: $(pf)(x) = f'(x)$, $(qf)(x) = xf(x)$. Its standard domain $\mathscr{S}(\mathbb{R})$ has to be restricted because of q^{-2} occurring in (4). We can take, e.g., the following common dense invariant domain of Q^B_+, Q^B_-, and Q^B_3:

$$D = \{\psi_{kn} \mid \psi_{kn}(x) = x^k \exp[-x^2 - (nx^2)^{-1}];$$

$$k = 0, \pm 1, \cdots, n = 1,2,\cdots\}_{\text{lin}}.$$

The term $-(nx^2)^{-1}$ in the exponent of ψ_{kn} ensures that D is in the domain of q^{-2}: $q^{-2}\psi_{kn} = \psi_{k-2,n}$, $\psi_{kn}(0) = 0$ for all k and n, and the ψ_{kn} are continuous on $\mathbb{R}$.

2.4. As mentioned above, we choose $\mathscr{H} = \mathscr{H}_B \otimes \mathscr{H}_F$ for the representation space, $\mathscr{H}_F$ being a "fermion" Hilbert space to be specified later. The expressions for the even generators are obtained by extending the boson representation (4) on $\mathscr{H}$ as follows:

$$\tilde{Q}_+ = Q^B_+ \otimes I_F,$$

$$\tilde{Q}_3 = Q^B_3 \otimes I_F, \quad \tilde{Q}_- = \tfrac{1}{4}i(p^2 \otimes I_F) + i(q^{-2} \otimes A). \tag{5}$$

The commutation relations (1a) are obviously satisfied for an arbitrary operator A; if we want these operators to be skew-symmetric A has to be symmetric.

2.5. The prescription (5) has to be completed by expressions for operators $\tilde{V}_+, \tilde{V}_-$ representing the odd generators. We shall adopt the following ansatz:

$$\tilde{V}_+ = q \otimes C_1, \quad \tilde{V}_- = q^{-1} \otimes C_2 + p \otimes C_3. \tag{6}$$

Then the commutation relations (1b) are fulfilled, as can be easily seen by comparing (1b) to the relations

$$[Q^B_3,q] = \tfrac{1}{2}q,$$

$$[Q^B_3,\alpha p + \beta q^{-1}] = -\tfrac{1}{2}(\alpha p + \beta q^{-1}), \quad \alpha,\beta \in \mathbb{C}$$

which immediately follow form Eqs. (3) and (4). For specifying the "fermion" operators C_r $(r = 1,2,3)$ and A, we make use of (1c); the remaining relations (1d) follow from (1a)–(1c) so that they provide no new conditions for the fermion operators. After simple calculations we get the following result:

Let the operators C_1, C_2 on $\mathscr{H}_F$ fulfill

$$C_1^2 = \tfrac{1}{4}iI_F, \quad [C_1,C_2]_+ = 0. \tag{7}$$

Then the operators

$$\tilde{Q}_+ = iq^2 \otimes I_F, \quad \tilde{Q}_3 = \tfrac{1}{2}(qp + \tfrac{1}{2}I_B) \otimes I_F,$$

$$\tilde{Q}_- = \tfrac{1}{4}ip^2 \otimes I_F + iq^{-2} \otimes (2C_1C_2 + 4iC_2^2), \tag{8}$$

$$\tilde{V}_+ = q \otimes C_1, \quad \tilde{V}_- = \tfrac{1}{2}ip \otimes C_1 + q^{-1} \otimes C_2,$$

satisfy the relations (1a)–(1d) and the Casimir operator is expressed as

$$K_2 = -\tfrac{1}{16}I - 4iI_B \otimes C_2^2. \tag{9}$$

2.6. Let us now examine under which conditions Eqs. (7),(8) determine an irreducible or at least Schurean representation of osp(1,2) supposing $\dim \mathscr{H}_F < \infty$.

Lemma 1: If the set $\{C_1,C_2\} \subset \mathrm{End}\mathscr{H}_F$ is reducible, then the operators (8) form a reducible set in $\mathrm{End}(\mathscr{H}_B \otimes \mathscr{H}_F)$.

Proof: This assertion immediately follows if one uses the $\sum_{i=1}^{n} \oplus \mathscr{H}_B$ realization of $\mathscr{H}_B \otimes \mathscr{H}_F$ $(\dim \mathscr{H}_F = n)$, in which $A \otimes B$ is expressed via matrix representation (b_{ij}) of B by

$$(A \otimes B)(x_1,\cdots,x_n) = \left(\sum_{j=1}^{n} b_{1j}Ax_j, \cdots, \sum_{j=1}^{n} b_{nj}Ax_j\right). \quad \blacksquare$$

Therefore, we shall consider only irreducible sets $\{C_1,C_2\}$ satisfying Eqs. (7). The Schur lemma implies that K_2 is then a multiple of unity, since $[C_1,C_2]_+ = 0$ implies $[C_1,C_2^2] = 0$, i.e., Schurean representations arise. The next lemma shows that such sets exist only if $\dim \mathscr{H}_F \leqslant 2$.

Lemma 2: (a) If $2 < \dim \mathscr{H}_F < \infty$, then any solution $\{C_1,C_2\}$ of Eqs. (7) is a reducible set. (b) If $\dim \mathscr{H}_F \leqslant 2$, then Eqs. (7) have the following solution only (up to similarity transformations):

$$\dim \mathscr{H}_F = 1, \quad C_1 = \pm \tfrac{1}{2}\varepsilon, \quad C_2 = 0, \tag{10}$$

$$\dim \mathscr{H}_F = 2, \quad C_1 = \tfrac{1}{2}\varepsilon\sigma_2, \quad C_2 = c\tfrac{1}{2}\varepsilon\sigma_1, \tag{11}$$

where $\varepsilon = \exp(i\pi/4)$, $c \in \mathbb{C}$, and σ_i are the Pauli matrices.

Proof: Let x be an eigenvector of C_2:

$$C_2x = c(\varepsilon/2)x, \quad c \in \mathbb{C},$$

and let $y = (2/\varepsilon)C_1x$. The linear envelope of $\{x,y\}$ is invariant with respect to both C_1 and C_2 so that the irreducibility condition implies $\{x,y\}_{\text{lin}} = \mathscr{H}_F$, i.e., $\dim \mathscr{H}_F = 1,2$. The remaining statements of the lemma can be easily verified combining the defining relations for x and y with Eqs. (7). $\blacksquare$

2.7. Consider the case $\dim \mathscr{H}_F = 1$. Substituting (10)

into (8) and (9) yields

$$\tilde{Q}_+ = iq^2, \quad \tilde{Q}_3 = \tfrac{1}{2}(qp + \tfrac{1}{2}I), \quad \tilde{Q}_- = i(p^2/4), \tag{12a}$$

$$\tilde{V}_+ = \pm\tfrac{1}{2}\varepsilon q, \quad \tilde{V}_- = \pm\tfrac{1}{2}i\varepsilon p, \tag{12b}$$

$$K_2 = -\tfrac{1}{16}I. \tag{12c}$$

We thus have two sets of operators $\tilde{Q}_\alpha, \tilde{V}_\beta$ differing in the sign of $\tilde{V}_\beta$. However, these sets are equivalent, the equivalence between them being established, e.g., by the unitary operator R on $L^2(\mathbb{R})$ which is given by $(R\psi)(x) = \psi(-x)$.

All the operators (12a) and (12b) are in the vector space End$D$ where D is the domain discussed in 2.3. In order to get a representation of osp(1,2) in terms of them, one has to introduce a Lie superalgebra structure in End$D$ such that the operators $\tilde{Q}_\alpha$ become even and $\tilde{V}_\beta$ odd elements.[13] To this purpose we use the linear manifolds D_0, D_1 of even and odd functions

$$D_k = \{\psi \in D \,|\, \psi(-x) = (-1)^k \psi(x)\}, \quad k = 0,1,$$

and by means of them introduce the following manifolds in End$D$:

$$\tilde{L}_j = \{T \in \mathrm{End}D \,|\, TD_k \subset D_{k+j(\mathrm{mod}\,2)}, k = 0,1\} \quad j = 0,1.$$

Then $D = D_0 \oplus D_1$, End$D = \tilde{L}_0 \oplus \tilde{L}_1$, and the operators (12a) are in $\tilde{L}_0$ (even) while (12b) in $\tilde{L}_1$ (odd). Hence a representation of osp(1,2) on $L^2(\mathbb{R})$ is determined by Eqs. (12a) and (12b). Each element of osp(1,2) is represented in terms of one pair of boson operators p,q. Therefore, this representation will be called "pure boson representation" (PBR).

Obviously, this representation has the ESS property. Notice that the PBR becomes a star representation of the complexification osp(1,2;$\mathbb{C}$), if one defines involution (adjoint operation) on osp(1,2;$\mathbb{C}$) by

$$Q_\alpha^* = -Q_\alpha, \quad V_\beta^* = -iV_\beta. \tag{13}$$

On the other hand, the PBR is not a grade star representation.[8] If it were so, then the relations between the standard involution " + " and the standard graded involution "#" on End$D$ [8] would imply $\tilde{Q}_\alpha^{\#} = \tilde{Q}_\alpha^{+} = -\tilde{Q}_\alpha$. As the PBR is faithful, a graded involution on osp(1,2) satisfying $Q_\alpha^{\#} = -Q_\alpha$ would have to exist. Then the axioms of graded involution and relations (1) would imply

$$V_\beta^{\#} = c_\beta V_\beta, \quad c_\beta \in \mathbb{C},$$

which contradicts the axiom $V_\beta^{\#\#} = -V_\beta$.

2.8. In the case $\dim\mathscr{H}_F = 2$ it is convenient to express the operators C_1, C_2 in (11) by means of

$$a = \tfrac{1}{2}(\sigma_1 - i\sigma_2), \quad a^+ = \tfrac{1}{2}(\sigma_1 + i\sigma_2).$$

The operators a, a^+ satisfy the CAR

$$a^2 = (a^+)^2 = 0, \quad [a, a^+]_+ = I_F,$$

and it further holds

$$[a^+, a] = \sigma_3.$$

Substituting (11) into Eqs. (8) and (9), we get

$$\tilde{Q}_+ = iq^2 \otimes I_F, \quad \tilde{Q}_3 = \tfrac{1}{2}(qp + \tfrac{1}{2}I_B) \otimes I_F,$$

$$\tilde{Q}_- = i(p^2/4 - c^2q^{-2}) \otimes I_F + \tfrac{1}{2}icq^{-2} \otimes [a^+, a], \tag{14a}$$

$$\tilde{V}_- = \tfrac{1}{4}\varepsilon p \otimes (a^+ - a) + \tfrac{1}{2}c\varepsilon q^{-1} \otimes (a + a^+),$$

$$\tilde{V}_+ = \tfrac{1}{2}i\varepsilon q \otimes (a - a^+), \tag{14b}$$

$$K_2 = (c^2 - \tfrac{1}{16})I. \tag{14c}$$

In order to embed the operators (14a) and (14b) into a Lie superalgebra, let us take the $(\mathscr{H}_B \oplus \mathscr{H}_B)$ realization of the tensor product $\mathscr{H}_B \otimes \mathscr{H}_F$ and the vector space End$D^{(2)}$, where

$$D^{(2)} = \{(x,y) \in \mathscr{H}_B \oplus \mathscr{H}_B \,|\, x,y \in D\}.$$

The Lie superalgebra structure End$D^{(2)} = \tilde{L}_0 \oplus \tilde{L}_1$ can be introduced as in 2.7; the only difference consists in replacing the linear manifolds D_k by $D_k^{(2)}$:

$$D_0^{(2)} = \{(x,0) | x \in D\}, \quad D_1^{(2)} = \{(0,y) | y \in D\}.$$

Then the operators (14a) and (14b) form a representation of osp(1,2); this assertion immediately follows from the fact that $A \otimes B \in \tilde{L}_0$ iff (b_{ij}) is a diagonal matrix and $A \otimes B \in \tilde{L}_1$ iff (b_{ij}) is antidiagonal (see the proof of Lemma 1 in 2.6).

The representation we have obtained is expressed in terms of one pair of boson operators p,q and one pair of fermion operators a, a^+, is Schurean, and depends on one complex parameter c; it will be called "fermion–boson representation" (FBR). For $c = 0$ the FBR is, of course, reducible [see Eq. (11)] and is equivalent to direct sum of two pure boson representations.

Has the FBR the ESS property? The standard involution " + " on End$D^{(2)}$ yields $\tilde{Q}_+^{+} = -\tilde{Q}_+, \tilde{Q}_3^{+} = -\tilde{Q}_3$, and $\tilde{Q}_-^{+} = -\tilde{Q}_- + iq^{-2}(\bar{c} - c) \otimes ((\bar{c} + c)I_F - \tfrac{1}{2}[a^+, a])$. Thus the ESS requirement is fulfilled iff $c \in \mathbb{R}$. For real c one further gets

$$\tilde{V}_\beta^{+} = -i\tilde{V}_\beta.$$

Hence the FBR is a star representation of osp(1,2;$\mathbb{C}$) with respect to the involution (13). On the other hand, the FBR is not a grade star representation for the same reasons as in the case of PBR.

III. CONCLUSIONS AND DISCUSSION

3.1. Our analysis of the representations of osp(1,2) arising from the boson representation (4) of sp(2,$\mathbb{R}$) can be summarized as follows:

(i) If $\dim\mathscr{H}_F > 2$, then there is no irreducible representation of the form (8); if $\dim\mathscr{H}_F \leqslant 2$, then any representation determined by Eqs. (7) and (8) is Schurean.

(ii) In the case $\dim\mathscr{H}_F = 1$ the Eqs. (7) and (8) determine just one representation (up to equivalence), namely the PBR [see (12a) and (12b)].

(iii) In the case $\dim\mathscr{H}_F = 2$ each representation determined by Eqs. (7) and (8) belongs (up to equivalence) to the family of FBR [see Eqs. (14a) and (14b)]. This family depends on one complex parameter c which is related to the Casimir operator by $K_2 = (c^2 - \frac{1}{16})I$.

(iv)The PBR is a star representation of osp(1,2;$\mathbb{C}$) with respect to the involution (13); the FBR is so iff $c \in \mathbb{R}$. In particular, the real forms of PBR and FBR have the ESS property (even elements are represented by skew-symmetric operators).

(v) Neither the PBR nor the FBR (for any $c \in \mathbb{C}$) is a grade star representation.

3.2. The pure boson representation can easily be generalized for the osp(1,2n) superalgebra. To this purpose one has to take n pairs of boson operators p_r, q_r obeying the CCR, p_r

being skew-symmetric and q_r symmetric. It is well known that the operators

$$X_{rs} = \tfrac{1}{2}[q_r,p_s]_+,$$
$$X_{r,-s} = -\tfrac{1}{2}i[q_r,q_s]_+ \equiv X_{s,-r},$$
$$X_{-r,s} = -\tfrac{1}{2}i[p_r,p_s]_+ \equiv X_{-s,r},$$

$r,s = 1,2,..n$, generate a skew-symmetric representation of the Lie algebra sp($2n$, $\mathbb{R}$). Let us further introduce

$$V_r = i\exp(i\pi/4)q_r, \quad V_{-r} = i\exp(-i\pi/4)p_r,$$
$$r = 1,2,..,n.$$

Then the following relations hold:

$$[X_{ij},V_k] = \delta_{kj}V_i - \epsilon_i\epsilon_j\delta_{-i,k}V_{-j},$$
$$[V_k,V_l]_+ = 2\epsilon_l X_{k,-l} = 2\epsilon_k X_{l,-k},$$

where $i, j, k, l = \pm 1, \pm 2,.., \pm n$ and $\epsilon_i = \mathrm{sgn}i$. Thus, the operators X_{ij},V_k generate a representation of osp(1,2n) which is ESS.

3.3. Canonical realizations of Lie algebras[11] provide an effective method for constructing infinite-dimensional representations. The possibility of applying an analogous procedure to Lie superalgebras seems to be promising: notice that the method is essentially an algebraic one. To this purpose one has to choose a suitable generalization of the Weyl algebra W_{2n}. We can take the most straightforward generalization[14,15] omitting other possibilities[16]: the associative algebra $W_{2n;m}$ with unity generated by the elements $q_i,p_i,i = 1,2..,n$ and ξ_k, $k = 1,2,..,m$, which obey the relations

$$[p_i,q_j] = \delta_{ij}\mathbf{1}, \quad [p_i,p_j] = [p_i,q_j] = 0,$$
$$[\xi_k,\xi_l]_+ = \delta_{kl}\mathbf{1}, \quad [p_i,\xi_k] = [q_i,\xi_k] = 0.$$

Notice that the (complex) algebra generated by n pairs of boson operators b_i,b_i^+ and m pairs of fermion operators a_k,a_k^+ is identical with $W_{2n;2m}$: it is sufficient to choose $\xi_k = 2^{-1/2}(a_k + a_k^+)$, $\xi_{m+k} = i2^{-1/2}(a_k - a_k^+)$. Then the problem we have solved in Sec. II may be formulated as searching for Schurean realizations of osp(1,2) in the Weyl superalgebra $W_{2;2}$.

3.4. It was pointed out in 1.4 that our approach does not work for each skew-symmetric boson representation of L_0 if we want to obtain an ESS representation of L. As an illustration let us consider the following example. The operators

$$Q_+ = -p, \quad Q_- = q^2p + q - i\gamma q,$$
$$Q_3 = -qp - \tfrac{1}{2}(1 - i\gamma)I_B \qquad (15)$$

generate for each $\gamma\in\mathbb{R}$ a skew-symmetric Schurean representation of sp(2,$\mathbb{R}$) on $\mathscr{H}_B = L^2(\mathbb{R})$. If we replace the parameter γ by a symmetric operator A on $\mathscr{H}_F$ ($\dim\mathscr{H}_F < \infty$), we get skew-symmetric operators on $\mathscr{H}_B \otimes \mathscr{H}_F$ which represent even generators of osp(1,2) [cf. Eq.(5)]:

$$\hat{Q}_+ = -p\otimes I_F, \quad \hat{Q}_- = (q^2p + q)\otimes I_F - iq\otimes A,$$
$$\hat{Q}_3 = -(qp + \tfrac{1}{2}I_B)\otimes I_F + \tfrac{1}{2}iI_B\otimes A.$$

For getting the operators $\hat{V}_+,\hat{V}_-$ we use now the relations

$$[\hat{Q}_+,\hat{V}_+] = 0, \quad [\hat{Q}_3,\hat{V}_+] = \tfrac{1}{2}\hat{V}_+,$$
$$[\hat{Q}_-,\hat{V}_+] = \hat{V}_-, \quad [\hat{Q}_-,\hat{V}_-] = 0 \qquad (16)$$

[the relations $[\hat{Q}_+,\hat{V}_-] = \hat{V}_+$, $[\hat{Q}_3,\hat{V}_-] = -\tfrac{1}{2}\hat{V}_-$ follow from (16) and hence provide no further conditions for $\hat{V}_+,\hat{V}_-$]. The first of the relations (16) implies $\hat{V}_+ = V(p)$, where

$$V(p)(x_1,...,x_n) = \left(\sum_{j=1}^n V_{1j}(p)x_j,...,\sum_{j=1}^n V_{nj}(p)x_j\right), \quad x_j\in D$$

(cf. the proof of Lemma 1 in 2.6), the operators $V_{kj}(p)\in\mathrm{End}D$ being functions of p only. Using the remaining three relations (16) and the fact that A is a symmetric operator on a finite-dimensional Hilbert space $\mathscr{H}_F$, one easily finds

$$\hat{V}_+ = \hat{V}_- = 0.$$

As this result contradicts anticommutation relations (1c), we conclude that the boson representation (15) cannot be extended to an ESS representation of osp(1,2).

3.5. It is obvious that our results remain valid also for the other real form of osp(1,2;$\mathbb{C}$) whose basis is obtained from Eqs. (1) replacing $V_\pm$ by $iV_\pm$.

3.6. Finally, let us remark that a pure boson representation akin to our PBR was obtained within a different framework in Ref. 17.

ACKNOWLEDGMENTS

The authors are obliged to Professor A. A. Kirillov, Dr. D. A. Leites, and Dr. V. I. Ogievetsky for discussions.

[1] V. G. Kac, Commun. Math. Phys. **53**, 31–64 (1977); V. G. Kac, Adv. Math. **26**, 8–96 (1977).
[2] A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062–73 (1975).
[3] D. Z. Djokovic, J. Pure Appl. Algebra **7**, 217–30 (1976).
[4] W. Nahm, V. Rittenberg, and M. Scheunert, Phys. Lett. B **61**, 383–84 (1976); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626–39, 1640–44 (1976).
[5] W. Nahm and M. Scheunert, J. Math. Phys. **17**, 868–79 (1976).
[6] P. Freund and I. Kaplansky, J. Math. Phys. **17**, 228–31 (1976).
[7] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155–62 (1977).
[8] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146–54 (1977).
[9] M. Bednář and V. Šachl, J. Math. Phys. **19**, 1487–92 (1978).
[10] M. Bednář and V. Šachl, J. Math. Phys. **20**, 367–71 (1979).
[11] P. Exner, M. Havlíček, and W. Lassner, Czech. J. Phys. B **26**, 1213–28 (1976).
[12] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN, Warsaw, 1977).
[13] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573–604 (1975).
[14] R. Casalbuoni, Nuovo Cimento A **33**, 115–25 and 389–431 (1976).
[15] F. A. Berezin and M. S. Marinov, Ann. Phys. **104**, 336–62 (1977).
[16] D. A. Leites, Dokl Akad. Nauk SSSR **236**, 804–7 (1977).
[17] R. Jagannathan and R. Vasudevan, J. Math. Phys. **20**, 390–2 (1979).

Lie-algebraic properties of infinite-dimensional wave equations

A. Cant
Department of Theoretical Physics, University of St. Andrews, St. Andrews, Fife KY16 9SS, Scotland

(Received 8 September 1980; accepted for publication 12 December 1980)

To an infinite-dimensional Lorentz-invariant wave equation of the form $(\alpha^\mu\partial_\mu + i\kappa)\psi(x) = 0$ is associated a Lie algebra S over $\mathbb{C}$ which contains so(4,$\mathbb{C}$) and α^μ. We show, by considering a certain class of equations, that in general S is an infinite-dimensional Lie algebra. It has a structure which is quite different from that of the known types of infinite-dimensional algebras.

PACS numbers: 02.20.Tw

I. INTRODUCTION

In previous papers[1-3] we have discussed the Lie-algebraic properties of finite-dimensional Lorentz-invariant wave equations of the form

$$\left(\alpha^\mu\frac{\partial}{\partial x^\mu} + i\kappa\right)\psi(x) = 0, \tag{1.1}$$

where the α^μ ($\mu = 0,1,2,3$) are $n\times n$ matrices and κ is a real nonzero constant. In this paper we make some comments on the Lie-algebraic structure of *infinite*-dimensional equations of the form (1.1).

We keep the notation of Refs. 1 and 2. We recall that a finite-dimensional equation of the form (1.1) is specified by a representation π of the Lorentz Lie algebra so (3,1) $\cong$sl $(2,\mathbb{C})^{\mathrm{R}}$ which admits a vector operator $\{\alpha^\mu\}$. We let S denote the Lie algebra generated by $\pi(D_2)$ and the α^μ over $\mathbb{C}$, where D_2 denotes the compexification so (4,$\mathbb{C}$) of so (3,1). Then, if we go over all irreducible representations ρ of S, we obtain a family of invariant wave equations based on the initial one.[1,2]

Furthermore, when there is a real form S_0 of S which contains sl$(2,\mathbb{C})^{\mathrm{R}}$, with the corresponding Lie group embedding SL$(2,\mathbb{C})\subset\mathscr{S}_0$, then we can include infinite-dimensional representations of $\mathscr{S}_0$ in this family.[3]

However, a new situation arises if we take an *infinite*-dimensional equation as our initial equation. We can define the Lie algebra S exactly as in the finite-dimensional case. After introducing the requisite notation in Sec. 2, we shall examine in Sec. 3 a particular class of infinite-dimensional equations, and demonstrate that S is almost always an infinite-dimensional Lie algebra. We conclude with some brief comments on the general structure of S.

2. NOTATION

Infinite-dimensional wave equations were discussed in detail by Gel'fand and Yaglom[4] in 1948; we shall use the notation of Ref. 5. The irreducible representations of sl $(2,\mathbb{C})^{\mathrm{R}}$ which are integrable to representations of the group SL (2,$\mathbb{C}$) are denoted by $\{l_0,l_1\}$, where l_0 is an integer or half-integer and $l_1\in\mathbb{C}$. The representation $\{l_0,l_1\}$ decomposes into irreducible representations of the maximal compact subalgebra su (2):

$$\{l_0,l_1\}\rightarrow \bigoplus_{j=|l_0|,|l_0|+1,\cdots}(j).$$

If $|l_1| > |l_0|$ and l_0,l_1 are simultaneously integral or half-integral, then $\{l_0,l_1\}$ is finite-dimensional, and the above sum terminates at $j = |l_1| - 1$. In all other cases $\{l_0,l_1\}$ is infinite-dimensional. Matrix elements of the generators of sl $(2,\mathbb{C})^{\mathrm{R}}$ in this representation can be found in Ref. 5.

An invariant wave equation is thus specified by a representation π which is the direct sum (or, more generally, direct integral) of representations of the form $\{l_0,l_1\}$. The vector operator α^μ can only have nonzero matrix elements between $\{l_0,l_1\}$ and $\{l_0',l_1'\}$ when

$$l_0' = l_0, \quad l_1' = l_1 \pm 1, \quad \text{or} \quad l_0' = l_0 \pm 1, \quad l_1' = l_1.$$

Explicit formulae for α^μ are given in Ref. 5.

3. THE STRUCTURE OF S IN THE INFINITE-DIMENSIONAL CASE

To get an idea of the properties of S for a given infinite-dimensional wave equation, we shall consider the following family of equations[6]:

$$\pi = \pi_1\oplus\pi_2 = \{\tfrac{1}{2},l_1\}\oplus\{-\tfrac{1}{2},l_1\}, \tag{3.1}$$

acting in the space $V = V_1\oplus V_2$, where $l_1\neq \pm 3/2, \pm 5/2,\cdots$, so V is infinite-dimensional.

The matrices α^μ are of the form

$$\alpha^\mu = \begin{pmatrix} 0 & X^\mu \\ Y^\mu & 0 \end{pmatrix},$$

in which X^μ, Y^μ are matrices, each determined up to multiplication by an arbitrary nonzero complex factor. (These factors are unimportant in what follows.)

Theorem 3.1: For the wave equations described by (3.1), the Lie algebra S over $\mathbb{C}$ which is generated by $\pi(D_2)$ and the α^μ is (a) isomorphic to sp $(4,\mathbb{C})\cong$so$(5,\mathbb{C})$ if $l_0 = 0$, (b) *infinite*-dimensional when $l_1\neq 0$.

Proof: It has been shown by Bracken[6] that, for a suitable choice of the complex factors appearing in α^μ,

$$[\alpha_\mu,\alpha_\nu] = I_{\mu\nu} - 2il_1\tilde{I}_{\mu\nu}\gamma_5, \tag{3.2}$$

where $I_{\mu\nu}$ are the generators of π,

$$\tilde{I}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}I^{\rho\sigma} \quad (\epsilon_{0123} = 1),$$

and $\gamma_5 = I_{V_1}\oplus(-I_{V_2})$, I_{V_1} and I_{V_2} being the identity operators on V_1 and V_2.

(a) If $l_1 = 0$, we have $[\alpha_\mu,\alpha_\nu] = I_{\mu\nu}$, which means that we have $S\cong$so$(5,\mathbb{C})\cong$sp$(4,\mathbb{C})$. This proves (a).

 0022-2488/82/030354-03$02.50 © 1982 American Institute of Physics

(b) From (3.2) we have

$$[\alpha_0,\alpha_1] = I_{01} - 2il_1 I_{23}\gamma_5,$$

$$[\alpha_1,\alpha_2] = I_{12} + 2il_1 I_{03}\gamma_5, \quad \text{etc.}$$

Thus if $l_1 \neq 0$, we see that S contains

$$I_{23}\gamma_5 = \frac{-[\alpha_0,\alpha_1] + I_{01}}{2il_1}, \quad I_{03}\gamma_5 = \frac{[\alpha_1,\alpha_2] + I_{12}}{2il_1}, \quad \text{etc.}$$

Clearly the matrices $I_{\mu\nu}, I_{\mu\nu}\gamma_5$ all belong to S, and therefore S contains

$$\Gamma_1(x) = \pi_1(x) \oplus 0, \quad \Gamma_2(x) = 0 \oplus \pi_2(x) \quad (\forall x \in D_2).$$

We put

$$\alpha_0 = \begin{pmatrix} 0 & X_0 \\ Y_0 & 0 \end{pmatrix},$$

where X_0, Y_0 are given in Ref. 5 (p. 276). Then, for $m,n \in \mathbb{Z}^+$ and $x,y \in D_2$ we have

$$[\mathrm{ad}\Gamma_1(x)]^m\,[\mathrm{ad}\Gamma_2(y)]^n\alpha_0 = \begin{pmatrix} 0 & (-1)^n[\pi_1(x)]^m X_0[\pi_2(y)]^n \\ \cdots & 0 \end{pmatrix}.$$

By taking suitable repeated commutators, it is clear by the Poincaré-Birkhoff-Witt theorem[7] that S will contain all matrices of the form

$$\begin{pmatrix} 0 & \pi_1(u)X_0\pi_2(u^1) \\ \cdots & 0 \end{pmatrix} \quad \forall u,u^1 \in \mathbb{U}^* = \mathbb{U}\setminus\{1\}.$$

Hence $\mathbb{U}$ denotes the universal enveloping algebra of D_2, with identity 1.

We recall now that if V_1, V_2 are vector spaces over $\mathbb{C}$, then the finite topology on $\mathrm{Hom}(V_2,V_1)$ is defined by stipulating that a basis for the topology is given by the open sets[8]

$$O(\xi_i;\eta_i) = \{A \in \mathrm{Hom}(V_2,V_1) \mid A\xi_i = \eta_i,\ i = 1,\ldots,m\},$$

where $\{\xi_1,\ldots,\xi_m\}$ is a linearly independent set of vectors in V_2 and $\{\eta_1,\ldots,\eta_m\}$ is any set of vectors in V_1.

Consider the set $\mathbb{D} = \pi_1(\mathbb{U}^*)X_0\pi_2(\mathbb{U}^*) \subseteq \mathrm{Hom}(V_2,V_1)$. We shall show that $\mathbb{D}$ is dense in $\mathrm{Hom}(V_2,V_1)$ in the finite topology; it then follows immediately that S is infinite-dimensional.

We use the following criterion (Ref. 8, p. 251): "a subset A of Hom (V_2,V_1) is dense $\Leftrightarrow$ for every set $\{\xi_1,\ldots,\xi_m\}$ of linearly independent vectors in V_2 and every set $\{\eta_1,\ldots\eta_m\}$ of vectors in V_1, $\exists A \in \mathbf{A}$ such that $A\xi_i = \eta_i$, $i = 1,\ldots,m$."

Suppose, then, that $\{\xi_1,\ldots\xi_m\}$ are linearly independent in V_2. Since $\pi_2(\mathbb{U}^*)$ is an irreducible algebra of linear transformations of a complex vector space of countable dimension, it follows from a result of Dixmier, proved in Ref. 9, that $\pi_2(\mathbb{U}^*)$ is algebraically completely irreducible, i.e., dense in End V_2. Thus we can choose $A_2 \in \pi_2(\mathbb{U}^*)$ such that $\xi_i = A_2\ \xi_i$, $i = 1,\ldots,m$. Then, since $\ker X_0 = \{0\}$[5], the set $\{X_0\ \xi_i\}$ is linearly independent in V_1. Again, choose $A_1 \in \pi_1(\mathbb{U}^*)$ such that $\eta_i = A_1(X_0\xi_i) = (A_1X_0A_2)\ \xi_i$. Since $A_1X_0\ A_2 \in \mathbb{D}$, it is clear that $\mathbb{D}$ is dense in $\mathrm{Hom}(V_2,V_1)$.□

The case (a) of this theorem expresses the fact that the (half-integral spin) Majorana equation, for which $\pi = \{\frac{1}{2},0\}$, is a member of the family of equations based on Dirac's equation. It is obtained by considering the ladder representation of sp $(4,\mathbb{R})$.[3,10] On the other hand, the case (b) of the theorem tells us that no wave equation of the form (3.1) can be a member of the family based on *any* finite-dimensional equation; it is a genuinely new equation. We could, in principle, produce new families of wave equations by finding the representations of S.

It is clear that for infinite-dimensional equations, the Lie algebra S will almost always be infinite-dimensional. This situation is directly analogous to that for finite-dimensional equations (see Sec. 3 of Ref. 1). There we used the same kind of irreducibility argument to show that S is almost always the whole of the appropriate algebra so(V), sp(V), or sl(V), with the corresponding wave equation being almost never obtainable from a lower-order equation.

We observe that the irreducible representation $\rho = \{l_0,l_1\}$, acting on W, is self-contragradient—a nondegenerate bilinear form with matrix B is induced on W by means of its structure as an su(2)-module

$$B = \bigoplus_{j = |l_0|,|l_0|+1,\cdots} B_j\ ,$$

where

$$(B_j)_{mm'} = (-1)^{j-m}\delta_{m,-m'}.$$

Clearly the generators $I_{\mu\nu}$ of ρ are all skew relative to B. In other words we have an embedding

$$D_2 \subseteq \mathrm{so}(W)[\mathrm{sp}(W)],$$

according as l_0 is integral (half-integral), where so (W) [sp(W)] denotes the Lie algebra of *column-finite* matrices which are skew relative to B.

This situation is similar to the finite-dimensional case[1,2]; in fact, an argument analogous to that of Theorem 3.3 in Ref. 1 shows that, if V admits a vector-operator α^μ, we can often ensure that $S \subseteq \mathrm{so}(V)[\mathrm{sp}(V)]$, if a bilinear form on V is suitably chosen [this is possible, in particular, for Eq. (3.1)].

The Lie algebra S is, however, unlike any of the infinite-dimensional Lie algebras normally considered in the literature. First, S is a finitely generated algebra of operators on a Hilbert space; but, since S contains unbounded operators (the α^μ are closed but unbounded[11]), it is not a "classical" Lie algebra in the sense of Ref. 12. Second, S has an obvious filtration $S_0 \subset S_1 \subset S_2 \subset \cdots$, where

$S_0 = \pi(D_2)$,

$S_1 =$ subspace spanned by $\pi(D_2)$ and the $\alpha^\mu,\cdots$

$S_k =$ subspace spanned by $\pi(D_2)$ and the commutators

$[\alpha^{\mu_1}[\alpha^{\mu_2}[\cdots[\alpha^{\mu_{r-1}},\alpha^{\mu_r}]]\cdots], \quad (r \leqslant k)]$.

It is clear that $[S_i,S_j] \subseteq S_{i+j}$ and $\cup_i S_i = S$. However S does not possess an obvious gradation compatible with this filtration. Thus S does not fit into the range of filtered algebras considered, for example, by Rudakov[13] and Guillemin.[14]

ACKNOWLEDGMENTS

The work described in this paper is part of the author's Ph.D. thesis (University of Adelaide, 1978). I should like to thank Professor C. A. Hurst for many stimulating conversa-

tions; I also thank Dr. A. L. Carey and Professor J. F. Cornwell for their helpful comments.

This work was supported by a Postgraduate Research Award from the Australian Government, and by a University of Adelaide Research Grant. I am also grateful to the Royal Commission for the Exhibition of 1851 for an Overseas Scholarship.

[1]A. Cant and C. A. Hurst, J. Aust. Math. Soc. B **20**, 446 (1978).
[2]A. Cant, J. Math. Phys. **22**, 870 (1981).
[3]A. Cant, J. Math. Phys. **22**, 878 (1981).
[4]I. M. Gel'fand and A. M. Yaglom, Zh. Eks. Teor. Fiz. **18**, 703, 1096, 1105 (1948).
[5]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, New York, 1963).
[6]A. J. Bracken, J. Phys. A **8**, 800 (1975).
[7]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer-Verlag, New York, 1972).
[8]N. Jacobson, *Lectures in Abstract Algebra, Vol. II- Linear Algebra* (Van Nostrand-Reinhold, New York, 1953).
[9]J. Lepowsky, Trans. Am. Math. Soc. **176**, 1 (1973).
[10]D. Tz. Stoyanov and I. T. Todorov, J. Math. Phys. **9**, 2146 (1968).
[11]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).
[12]P. de la Harpe, *Classical Banach-Lie Algebras and Banach-Lie Groups of Operators in Hilbert Space*, Lecture Notes in Mathomatics, **285** (Springer-Verlag, Berlin, 1972).
[13]A. N. Rudakov, Math. USSR Izv. **8**, 836 (1974).
[14]V. Guillemin, J. Diff. Geom. **4**, 257 (1970).

Remarks on certain dual series equations involving the Konhauser biorthogonal polynomials [a)]

H. M. Srivastava
Department of Mathematics, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada

(Received 15 December 1980; accepted for publication 23 October 1981)

It is observed, in the present note, that the literature contains erroneous results concerning the solutions of certain dual (and triple) equations involving series of the Konhauser biorthogonal polynomials. For example, the main results proved recently by K. R. Patil and N. K Thakare [J. Math. Phys. **18**, 1724 (1977)] are shown to be invalid except in their already-known special cases. The errors are traced to the misuse of a certain Weyl fractional integral which holds true only in the case of the classical Laguerre polynomials.

PACS numbers: 02.30. − f, 02.30.Jm

Patil and Thakare (Ref. 1, p. 1725) claim to have solved some dual series equations involving the Konhauser biorthogonal polynomial set $\{Z_n^{\alpha}(x;k)\}_{n=0}^{\infty}$. Their method uses Abel's integral equations and is based rather heavily upon the following Weyl fractional integral of order β [Ref. 1, p. 1725, Eq. (2.2)]:

$$\int_{\xi}^{\infty} \exp(-x)(x-\xi)^{\beta-1} Z_n^{\delta+\beta}(x;k)\,dx = \Gamma(\beta)\exp(-\xi) Z_n^{\delta}(\xi;k), \quad \delta+1>\beta>0, \tag{1}$$

which, upon setting $x=\xi+t$, yields its equivalent form

$$Z_n^{\delta}(\xi;k) = \frac{1}{\Gamma(\beta)}\int_0^{\infty} \exp(-t)t^{\beta-1} Z_n^{\delta+\beta}(\xi+t;k)\,dt, \tag{2}$$

where, as before, $\delta+1>\beta>0$. For $k=1$, both (1) and (2) reduce to the corresponding well-known results for the Laguerre polynomials $L_n^{(\alpha)}(x)\equiv Z_n^{\alpha}(x;1)$.

In view of the definition of $Z_n^{\alpha}(x;k)$, the left-hand side of (2) is a polynomial of degree n in ξ^k, whereas the right-hand side is a polynomial also, of degree nk in ξ, in which all powers of ξ from 0 to nk appear with nonzero coefficients. Thus it would suffice to check the validity of (2) for $n=1$, in which case (2) readily yields

$$\frac{\Gamma(\delta+k+1)}{\Gamma(\delta+1)} - \xi^k = \frac{\Gamma(\delta+\beta+k+1)}{\Gamma(\delta+\beta+1)} - (-1)^k k!\, L_k^{(-\beta-k)}(\xi), \tag{3}$$

which (upon comparing the coefficients of like powers of ξ) is seen to hold true if and only if $k=1$.

It follows that (2), and hence also its equivalent form (1), are invalid for $k=2,3,4,\cdots$. Therefore, all of Sec. 3 of the Patil–Thakare paper,[1] except possibly their equation (3.1), would hold true only for $k=1$, that is, for the classical Laguerre polynomials. Indeed, such dual Laguerre series equations were solved earlier by Lowndes[2] and their various generalizations were considered subsequently by Srivastava (Refs. 3–5), and Srivastava and Panda.[6]

The integral formula (1) and its further erroneous versions have also appeared elsewhere in the literature. For example, the recent papers by Karande and Thakare [Ref. 7, p. 643, Eq. (3.6)] and Kumbhat [Ref. 8, p. 140, Eq. (2.4)] may be cited. The former attribute (1), by an obvious oversight, to Prabhakar,[9] while the latter uses (1), with a further error, in solving certain dual and triple series equations involving the Konhauser biorthogonal polynomials $Z_n^{\alpha}(x;k)$. The corrections to all these papers, and possibly many more using (1), would simply amount to viewing each of their results in the already-known context of the classical Laguerre polynomials.

[a)]Supported, in part, by NSERC Grant A7353.

[1]K. R. Patil and N. K. Thakare, J. Math. Phys. **18**, 1724 (1977).
[2]J. S. Lowndes, Pacific J. Math. **25**, 123 (1968).
[3]H. M. Srivastava, Pacific J. Math. **30**, 525 (1969).
[4]H. M. Srivastava, J. Math. Anal. Appl. **31**, 587 (1970).
[5]H. M. Srivastava, Nederl. Akad. Wetensch. Proc. Ser. A **76** = Indag. Math. **35**, 137 (1973).
[6]H. M. Srivastava and R. Panda, Nederl. Akad. Wetensch. Proc. Ser. A **81** = Indag. Math. **40**, 502 (1978).
[7]B. K. Karande and N. K. Thakare, Indian J. Pure Appl. Math. **7**, 635 (1976).
[8]R. K. Kumbhat, Vijnana Parishad Anusandhan Patrika **22**, 139 (1979).
[9]T. R. Prabhakar, Pacific J. Math. **35**, 213 (1970).

 0022-2488/82/030357-01$02.50 © 1982 American Institute of Physics

The bi-Hamiltonian structure of some nonlinear fifth- and seventh-order differential equations and recursion formulas for their symmetries and conserved covariants

Benno Fuchssteiner and Walter Oevel
Fachbereich 17 Mathematik-Informatik, Universität-Gesamthochschule-Paderborn, D 4790 Paderborn West Germany

(Received 24 December 1980; accepted for publication 4 September 1981)

Using a bi-Hamiltonian formulation we give explicit formulas for the conserved quantities and infinitesimal generators of symmetries for some nonlinear fifth- and seventh-order nonlinear partial differential equations; among them, the Caudrey–Dodd–Gibbon–Sawada–Kotera equation and the Kupershmidt equation. We show that the Lie algebras of the symmetry groups of these equations are of a very special form: Among the C^∞ vector fields they are generated from two given commuting vector fields by a recursive application of a single operator. Furthermore, for some higher order equations, those multisoliton solutions, which for $|t| \to \infty$ asymptotically decompose into traveling wave solutions, are characterized as eigenvector decompositions of certain operators.

PACS numbers: 02.30.Jr

I. THE MAIN RESULTS

It has long been known[1–6] that the Caudrey-Dodd-Gibbon-Sawada–Kotera (CDGSK) equation

$$u_t = K_1(u) = u_{xxxxx} + \tfrac{5}{2}\zeta u u_{xxx} + \tfrac{5}{2}\zeta u_x u_{xx} + \tfrac{5}{4}\zeta^2 u^2 u_x \quad (\zeta \text{ arbitrary} \in R) \tag{1}$$

is "completely integrable" in the sense that it admits infinitely many conservation laws and infinitely many symmetries (via a suitable version of Noether's theorem). Reference 1 provides explicit formulas for only four conserved quantities "because of the enormity of the calculation."

Satsuma and Kaup[2] have found a Bäcklund transformation and an inverse scattering method for (1) that enabled them to give an explicit recursion formula for the conserved densities. It turned out that many of those densities are trivial in the sense that the conserved charges vanish. Moreover, conserved densities of certain order in u do not seem to exist at all, e.g., there obviously is no conserved polynomial density with $u^2, u^5, u^8 \ldots$ as the highest power of u.

A. Our main result is that all symmetries and all nontrivial conserved densities of (1) can be obtained by a straightforward recursion scheme. We claim that (J_1, K_1, Θ_1) is a generalized bi-Hamilton system, where Θ_1 and J_1 are given by

$$\Theta_1(u) = D^3 + \zeta(uD + Du), \tag{2}$$

$$J_1(u) = 2D^3 + \zeta(D^2 u D^{-1} + D^{-1} u D^2) + \tfrac{1}{4}\zeta^2(u^2 D^{-1} + D^{-1} u^2). \tag{3}$$

Here D stands for the differential operator and D^{-1} for its inverse (defined on a suitable solution space that is explained later on). As will be shown in Sec. II the bi-Hamiltonian structure means that $\Theta_1(u)$ maps conserved convariants of (1) (especially the gradients of conservation laws) onto infinitesimal generators of one-parameter symmetry groups, and $J_1(u)$ works in the opposite way, i.e., infinitesimal generators of symmetries are mapped onto conserved covariants. Hence, the operator

$$\Phi_1(u) = \Theta_1(u) J_1(u) \tag{4a}$$

is a recursion operator in the sense that it maps infinitesimal generators of symmetries onto infinitesimal generators of symmetries, and

$$\Phi_1^+(u) = J_1(u)\Theta_1(u) \tag{4b}$$

maps conserved covariants of (1) onto conserved covariants.

We review some basic notions.

A vector field $\sigma(u)$ is said to be the *infinitesimal generator of a symmetry* of

$$u_t = K(u) \tag{5a}$$

if the infinitesimal transformation $u(t) \to u(t) + \epsilon\sigma(u(t))$ leaves (5a) form-invariant. In other words $v(t) = \sigma(u(t))$ has to be a solution of

$$v_t = \frac{\partial}{\partial \epsilon} K(u + \epsilon v)_{|\epsilon = 0}. \tag{5b}$$

A covector field $G(u)$ is called a conserved covariant if $\langle G(u(t)), v(t)\rangle$ is time independent whenever $u(t)$ is a solution of (5a) and $v(t)$ is a solution of (5b). Special conserved covariants are given by gradients of conservation laws. The quantity $p(u)$ is a conservation law of (5a) if $p(u(t))$ is time independent for every solution $u(t)$ of (5a). The gradient of $p(u)$ is defined by

$$\langle \operatorname{grad} p(u), v\rangle = \frac{\partial}{\partial \epsilon}\bigg|_{\epsilon = 0} p(u + \epsilon v).$$

Here $\langle\ ,\ \rangle$ stands for the bilinear form given by

$$\langle f, g\rangle = \int_R f(x) g(x)\, dx.$$

Now, we return to Eq. (1). Picking $\sigma_1(u) = u_x$ and $\sigma_3(u) = K_1(u)$ as the infinitesimal generators of space and time translation, we can construct two sequences of symmetry generators by

$$\sigma_{3n+1}(u) := \phi_1^n(u) u_x \quad \text{and} \quad \sigma_{3n+3}(u) := \phi_1^n(u) K_1(u). \tag{6a}$$

 0022-2488/82/030358-06$02.50 © 1982 American Institute of Physics

The subscripts are chosen in such a way that they indicate the order in u (i.e., the maximal power of u) of each generator. Analogously there are 2 hierarchies of conserved covariants that are connected to the symmetry generators via Θ_1 and J_1 and can as well be constructed from simple ancestors. The simplest conserved densities of (1) are u and $[(\zeta/12)u^3 - u_x^2/2)]$ so that we pick their gradients as the starting points of the two hierarchies. Defining

$$G_0(u) = \operatorname{grad} \int_{-\infty}^{\infty} u\,dx = 1$$

and

$$G_2(u) = \operatorname{grad} \int_{-\infty}^{+\infty} \left(\frac{\zeta}{12} u^3 - \frac{u_x^2}{2} \right) dx = \frac{\zeta}{4} u^2 + u_{xx},$$

we obtain the sequences

$$G_{3n} = \phi_1^+(u)^n 1 \quad \text{and} \quad G_{3n+2} = \phi_1^+(u)^n \left(\frac{\zeta}{4} u^2 + u_{xx} \right) \tag{6b}$$

for which the relations

$$\sigma_{k+1}(u) = \Theta_1(u) G_k(u) \quad \text{and} \quad J_1(u)\sigma_k(u) = G_{k+2}(u) \tag{7}$$

hold for $k = 0,2,3,5...$ and $k = 1,3,4,6...$, respectively.

In the particular case under consideration all the quantities $G_k(u)$ are gradients. Hence the corresponding conservation laws can be found by integration, i.e., all the

$$I_{k+1}(u) = \int_0^1 \langle G_k(\lambda u), u \rangle\, d\lambda \tag{8}$$

are conserved densities. The first four of these densities can be found in Ref. 1.

Looking at the orders of the I_k we see that there are the two hierarchies I_{3n+1} and I_{3n+3}, so we have verified the conjecture "that a series of conservation laws exists with every third polynomial conservation law (p.c.l.) missing, i.e., if $\frac{1}{3}(2r-1)$ is an integer then the p.c.l. (of that order in u) does not exist."[1] This conjecture was already proved[6] (by the inverse scattering method and under strong compact-support assumptions). It should be noted that our results do not depend on inverse scattering techniques.

B. We shall see that the same analysis holds for the Kupershmidt equation[5] (among others):

$$u_t = K_2(u) = u_{xxxxx} + \tfrac{5}{2}\zeta u_{xxx} u + \tfrac{25}{4}\zeta u_{xx} u_x + \tfrac{5}{4}\zeta^2 u^2 u_x. \tag{9}$$

Here, the crucial operators $\Theta(u)$ and $J(u)$ have the following form:

$$\Theta_2(u) = D^3 + \tfrac{1}{4}\zeta(uD + Du), \tag{10}$$

$$J_2(u) = 2D^3 + \tfrac{3}{2}\zeta(uD + Du) + \zeta(D^2 u D^{-1} + D^{-1} u D^2) + \zeta^2(u^2 D^{-1} + D^{-1} u^2), \tag{11}$$

and the two gradients of conservation laws we have to start with are in this case

$$G_0(u) = 1, \tag{12a}$$

$$G_1(u) = u_{xx} + \zeta u^2 = \operatorname{grad} \int_{-\infty}^{+\infty} \left(\frac{\zeta}{3} u^3 - \frac{u_x^2}{2} \right) dx. \tag{12b}$$

C. Another fifth-order equation that can be dealt with in the same way is

$$v_t = v_{xxxxx} - 5(v_x v_{xxx} + v_{xx}^2 + v_x^3 + 4v v_x v_{xx} + v^2 v_{xxx} - v^4 v_x). \tag{13}$$

It was recently discovered[5] that this equation is connected to (1) as well as to (9) by a Bäcklund transformation (modified Miura transformation). Hence we can apply the transformation formulas of Ref. 7 to obtain the corresponding operators $\tilde{\Theta}(v), \tilde{J}(v)$, as well as the conserved covariants to start our two sequences.

Equation (1) goes over into (13) via

$$v_x - v^2 = \tfrac{1}{2}\zeta u. \tag{14}$$

In Sec. III we explain that such a formula, which is called a Bäcklund transformation, yields transformation formulas[7,13] for the corresponding operators Θ and J. By application of these transformation formulas we obtain the corresponding operators $\tilde{\Theta}(v)$ and $\tilde{J}(v)$ for Eq. (13). They are given by

$$\tilde{\Theta}(v) = \frac{4}{\zeta^2}(2v - D)^{-1} \Theta_1\left(\frac{2}{\zeta}(v_x - v^2) \right)(2v + D)^{-1}, \tag{15a}$$

$$\tilde{J}(v) = \frac{\zeta^2}{4}(2v + D) J_1\left(\frac{2}{\zeta}(v_x - v^2) \right)(2v - D), \tag{15b}$$

where $J_1(\cdot)$ and $\Theta_1(\cdot)$ are the operators given by (2) and (3). The conserved covariants to obtain the two sequences of conservation laws from are

$$G_0(v) = 1, \tag{16}$$

$$G_2(v) = (D + 2v)[(v_x - v^2)^2 + 2(v_x - v^2)_{xx}]. \tag{17}$$

D. In addition our method yields recursion formulas for conservation laws and infinitesimal generators of symmetries of some seventh-order equations that were not yet discovered as completely integrable. To be precise, take any pair of operators Θ, J given by either (2), (3) or (10), (11) or (15a), (15b). Then the equations

$$u_t = \Phi(u)^n u_x, \quad \text{where} \quad \Phi(u) = \Theta(u) J(u), \tag{18}$$

are completely integrable and the operators Θ, J play the same role as before.

The infinitesimal generators of one-parameter symmetry groups are given by

$$\sigma_n(u) = \Phi(u)^n u_x, \quad n \in N, \tag{19a}$$

and

$$G_n(u) = \Phi^+(u)^n G_0(u), \quad n \in N, \tag{19b}$$

[where $G_0(u) = 1$ and $\Phi^+(u) = J(u)\Theta(u)$] are gradients of conserved quantities. Furthermore for Eq. (18) we have a very transparent description of the multisoliton solutions. As described in Sec. III (see also Refs. 8 and 9) a solution u of (18) is a pure N-soliton, i.e., a solution that decomposes for $|t| \to \infty$ into N traveling wave solutions with prescribed as-

ymptotic speeds such that the total energy is carried by the asymptotic waves, if and only if u_x can be written as

$$u_x = \sum_{k=1}^{N} \omega_k, \tag{20a}$$

where

$$\Phi(u)\omega_k = -\lambda_k \omega_k, \quad k = 1,...,N, \tag{20b}$$

i.e., the ω_k are eigenvectors of $\Phi(u)$. Hence the manifold of N-soliton solutions of (18) can be described in terms of the solution manifold of a system of ordinary differential equations. The asymptotic speeds are given by the eigenvalues, and are equal to $\lambda^n_1,...,\lambda^n_N$.

II. COMPARISON WITH OTHER WORK

A recursion formula for the conserved quantities (integrals) of the CDGSK equation has been given before (see, for example, Ref. 2). This recursion formula is rather complicated since the nth integral I_n is a polynomial of third order in the I_k, $k < n$. Although this formula is the result of ingenious considerations, it seems absolutely hopeless to find an explicit solution for this recursion scheme. But indeed, formula (6b) constitutes such an explicit solution. To be precise, it constitutes an explicit formula for the gradients of the integrals. But obtaining the integrals from their gradients is only an elementary calculation [formula (8)]. In fact, dealing with the gradients instead of the integrals has a considerable advantage since one immediately works within the framework of the Poisson brackets. In addition, the operator Φ^+ [given by formula (4b)] is a deformation in that Lie-algebra of Poisson brackets having strong linear interpolation properties[14]—and since the kernel of Φ^+ is empty, one can be sure to construct only those integrals for which the corresponding densities do not vanish.

We should like to emphasize that our results are not obtained by inverse scattering techniques. Although we do appreciate inverse scattering as one of the most ingenious contributions in the field, we cannot overlook the fact that, for the equations under consideration, the validity of the results obtained by this method depends, up to now, on very strong compact support assumptions (see Ref. 6, p. 215). In contrast, all the results obtained in this paper can be checked by direct (although cumbersome) calculations.

Our approach yields in addition a complete description of the symmetry group of the CDGSK equation (in terms of infinitesimal generators). Symmetries for this equation have not been considered in detail. Of course, barring some minor complications, one obtains the symmetry group from the conserved quantities via a suitable version of Noether's law. But the point is, that for finding such a version of Noether's law one has to perform, at least partly, those calculations that are presented in this paper (namely, finding Hamiltonian formulations for the equations under consideration). Since the infinitesimal generators of the one-parameter symmetry groups do commute, one has, in addition, found the generalized CDGSK equations. Here "generalized" is meant in exactly the same sense as in the case of the Korteweg–de-Vries (KdV) equation.

III. THE BACKGROUND

We consider a dynamical system

$$u_t = K(u), \tag{21}$$

where u is in some manifold M and where K is a suitable C^∞ vector field. For our purpose we assume throughout this paper that $M = S =$ the Schwartz space of C^∞ functions on the real line vanishing rapidly at $\pm\infty$. Because of this assumption we can identify the manifold with the typical fiber of the tangent bundle, and everything becomes very transparent. In the C^∞ vector fields we introduce the usual Lie-algebra product (see any standard textbook, for example Ref. 10 or 11) given by

$$[H_1,H_2](u) = \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} [H_1(u + \epsilon H_2(u)) - H_2(u + \epsilon H_1(u))],$$

where H_1, H_2 are arbitrary C^∞ vector fields.

The importance of this Lie algebra lies in the fact that a flow defined by a vector field H is a one-parameter symmetry group of (21) if and only if $[K, H] = 0$. Therefore if $[K, \sigma] = 0$, then σ is the *infinitesimal generator* of a *symmetry* and has the properties required in Sec. I.

Now, let us turn our attention towards the dual. Let S^* be a space of C^∞ functions on the real line such that the elements $f \in S^*$ define continuous linear functionals on S via

$$\langle f,s\rangle = \int_{-\infty}^{+\infty} f(\xi)\, s(\xi)\, d\xi, \quad s \in S.$$

Let us henceforth restrict all notions of continuity and differentiability to the topology given by the pair (S^*,S) (where S^* will be chosen later on in such a way that the operators we deal with do make sense). A C^∞ covector field G is called a *conserved covariant* if $L_K G = 0$, where L_K denotes the Lie derivative. That is the same as saying that G defines a flow in the cotangent bundle such that $\langle G, K\rangle$ is a conserved quantity for (21). Gradients of conservation laws are standard examples for conserved quantities; in that case we have trivially $\langle G,K\rangle = 0$. But in general a conserved covariant need not be closed.

It is very useful to look for a relationship between symmetries and conserved quantities of the dynamical system given by $K(u)$. Such a connection similar to Noether's theorem can be formulated, if $K(u)$ has a special, i.e., Hamiltonian, structure. To make precise what this means, we first have to establish the notion of symplectic and implectic (inverse-symplectic) operators.

Definition: An operator-valued function $u \to J(u) : S \to S^*$ is called *symplectic* if (1) $J(u)$ is antisymmetric with respect to $\langle\ ,\ \rangle$; and (2) the 2-form ω defined by $\omega(s_1,s_2) := \langle J(u)s_1,s_2\rangle\ \forall s_1,s_2 \in S$ is closed.
A 2-form ω is closed if and only if the following Jacobi identity is satisfied:

$$L_{X_3}(\omega(X_1,X_2)) + L_{X_1}(\omega(X_2,X_3)) + L_{X_2}(\omega(X_3,X_1)) = 0$$

for all vector fields X_1, X_2, X_3. This is equivalent to

$$\langle J'(u)[s_1]s_2,s_3\rangle + \langle J'(u)[s_2]s_3,s_1\rangle + \langle J'(u)[s_3]s_1,s_2\rangle = 0 \quad \text{for all } s_1,s_2,s_3 \in S.$$

Here again L_X is the Lie-derivative with respect to the vector field X and $J'(u)[v]$ stands for the derivative

$$J'(u)[v] = \frac{\partial}{\partial \epsilon} J(u + \epsilon v)_{|\epsilon = 0}.$$

Analogously, an antisymmetric operator $\Theta(u) : S^* \to S$ is called implectic if it satisfies the same Jacobi identity as the inverse of a nondegenerate symplectic operator.

Definition: An operator $\Theta(u) : S^* \to S$ is called *implectic*, if (1) $\Theta(u)$ is antisymmetric with respect to $\langle\ ,\ \rangle$; and (2) the 2 times contravariant tensor field φ defined by

$$\varphi(X_1^*, X_2^*) := \langle X_1^*, \Theta(u) X_2^* \rangle \quad \forall X_1^*, X_2^* \in S^*$$

satisfies

$$L_{\Theta X_3^*} \varphi(X_1^*, X_2^*) + L_{\Theta X_2^*} \varphi(X_3^*, X_1^*) + L_{\Theta X_1^*} \varphi(X_2^*, X_3^*) = 0$$

for all covector fields X_1^*, X_2^*, X_3^*.

Again, this last identity is equivalent to

$$\langle s_2^*, \Theta'(u)[\Theta(u) s_1^*] s_3^* \rangle + \langle s_1^*, \Theta'(u)[\Theta(u) s_3^*] s_2^* \rangle + \langle s_3^*, \Theta'(u)[\Theta(u) s_2^*] s_1^* \rangle = 0 \quad \text{for all } s_1^*, s_2^*, s_3^* \in S^*.$$

Now we can characterize those dynamical systems $u_t = K(u)$ for which intimate relations between the symmetries and the conserved quantities exist. Let d denote the exterior derivative.

Definition: If there is a symplectic $\Theta(u) : S^* \to S$ and a function $H(\cdot) : S \to \mathbb{R}$ (i.e., a zero form) such that

$$K(u) = \Theta(u)\, dH(u)$$

then we call (K, Θ) a *generalized Hamiltonian system*.

If there is an implectic $J(u) : S \to S^*$ and a function $H(\cdot) : S \to \mathbb{R}$ such that

$$J(u) K(u) = dH(u),$$

then we call (J, K) *a generalized inverse Hamiltonian system*. Here "generalized" refers to the fact that we do not assume any nondegeneracy or invertibility conditions for the operators $\Theta(u)$ and $J(u)$.

The nice properties of these systems can be stated in a kind of generalized Noether Theorem:

Theorem:

(i) Let (K, Θ) be generalized *Hamiltonian*. Then Θ maps conserved covariant forms onto infinitesimal generators of symmetries.

(ii) Let (J, K) be generalized inverse Hamiltonian. Then J maps infinitesimal generators of symmetries onto conserved covariant forms. The proof of this theorem can be found in Ref. 13 and a detailed analysis of the Lie-algebraic aspects is contained in Ref. 12.

A most convenient case is a dynamical system that turns out to be generalized *bi-Hamiltonian*, i.e., there is a symplectic Θ and an implectic J such that (K, Θ) is generalized Hamiltonian and (J, K) is generalized inverse Hamiltonian.

For such a system (J, K, Θ) our theorem immediately yields that $\Phi(u) := \Theta(u) \circ J(u)$ and $\Phi^+(u) := J(u) \circ \Theta(u)$ are self-maps in the spaces of infinitesimal generators of symmetries and of conserved covariant forms, respectively. Therefore the bi-Hamiltonian structure of a system gives us a recursion operator in the sense of Olver,[15] which maps one generator of a symmetry to another one and the adjoint of which maps one conserved quantity to the next.

If this recursion process is not cyclic or stops at a certain stage, we can construct an infinite hierarchy of symmetries starting with simple ones such as space or time translation. To the members of this hierarchy correspond conserved covariant forms, which as well can be constructed via Φ^+ from common ancestors. This method goes back to an ingenious idea of Magri,[18] who considered bi-Hamiltonian systems under the additional hypothesis that Φ^+ is a self-map in those closed covector fields commuting (in the Lie algebra of Poisson brackets) with $J(u)K(u)$.

We claim that, for all the cases considered in Sec. I, the system (J, K, Θ) is bi-Hamiltonian; this then proves all assertions of Secs. IA and IB except the statement that the G_n are closed and in involution. The necessary calculations to prove that (J, K, Θ) is bi-Hamiltonian are given in Sec. IV; the reasons for the latter statement, and the claims of Sec. ID are presented in Sec. V.

Let us turn our attention to Sec. IC. Let $F : S \to S$ be a C^∞-diffeomorphism, mapping solutions of $v_t = H(v)$ onto solutions of $u_t = K(u)$. For example

$$u = (2/\zeta)(v_x - v^2).$$

Such a map is said to be a *Bäcklund transformation* between the two evolution equations. It is well known (Ref. 11, p. 132) that the derivative F' of a C^∞-diffeomorphism defines a Lie-algebra isomorphism in the vector fields; its adjoint F'^+ is an isomorphism with respect to the Poisson brackets. Application of these facts yields the formulas (15). (For more examples of such transformations see Refs. 7 and 13.)

IV. THE CALCULATIONS

Consider the vector field K given by

$$K(u) = u_{xxxxx} + \alpha_1 u u_{xxx} + \alpha_2 u_x u_{xx} + \alpha_3 u^2 u_x, \tag{22}$$

where $u \in S$ and $\alpha_1, \alpha_2, \alpha_3$ are numbers.

For $\alpha_3 = \frac{1}{10}(2\alpha_2 - \alpha_1)(3\alpha_1 - \alpha_2)$ this can be written as[13]

$$K(u) = \Theta(u)\, \text{grad} \int_{-\infty}^{+\infty} \left(-\frac{u_x^2}{2} + \frac{2\alpha_2 - \alpha_1}{30} u^3 \right) dx, \tag{23}$$

where

$$\Theta(u) = D^3 + \tfrac{1}{5}(3\alpha_1 - \alpha_2)(uD + Du). \tag{24}$$

The operator Θ is in fact implectic (already observed in Refs. 7 and 13, an easy calculation). Now, we consider the operator

$$J(u) = 2D^3 + \gamma(uD + Du) + \alpha(D^2 u D^{-1} + D^{-1} u D^2) + \beta(u^2 D^{-1} + D^{-1} u^2) + \delta u D^{-1} u, \quad u \in S \tag{25}$$

and we fix our dual space S^* in such a way that $J(u) : S \to S^*$ (which can be done easily). One proves that the corresponding 2-form is closed; hence $J(u)$ is symplectic. Now we want to determine our coefficients in such a way that $J(u)\, K(u)$ is a closed form (i.e., a gradient). This appears to be a simple exercise, but in fact is not (at least if the calculation is done by hand; some of our students are at present writing a computer program for calculations of this kind). Here, we only present the essential steps of that calculation.

One looks at different powers in u of $J(u)\,K(u)$. The first- and fifth-order terms are easily seen to be gradients.

Second order in u: The gradient of

$$A\int_{-\infty}^{+\infty} u_{xx}^3\,dx + B\int_{-\infty}^{+\infty} uu_{xx}^2\,dx$$

is equal to

$$(6A-B)u_{xxx}u_{xxx} + (6A-6B)u_{xx}u_{xxxx} - 6Bu_x u_{xxxxx} + 2Buu_{xxxxxx}.$$

If one assumes that the second-order term of $J(u)\,K(u)$ is of that form one obtains the following expressions:

$$\begin{aligned}
&\alpha = \tfrac{2}{5}\alpha_1,\\
&\gamma = \tfrac{2}{5}(\alpha_2-\alpha_1),\\
&A = \tfrac{2}{15}(\alpha_1+7\alpha_2),\\
&B = -\alpha_1-\tfrac{2}{5}\alpha_2.
\end{aligned}\tag{26}$$

Third order in u: We assume that the third order term is equal to the gradient of

$$C\int_{-\infty}^{+\infty} u_x^4\,dx + E\int_{-\infty}^{+\infty} u^2u_{xx}^2\,dx.$$

This leads to the following equations for the coefficients:

$$\begin{aligned}
&\alpha_3 = \tfrac{1}{10}(3\alpha_1-\alpha_2)(2\alpha_2-\alpha_1),\\
&\beta = \tfrac{1}{25}\alpha_1(2\alpha_2-\alpha_1),\\
&\delta = \tfrac{4}{25}\left|\alpha_2-\tfrac{5}{2}\alpha_1\right|(\alpha_1-\alpha_2),\\
&C = \tfrac{1}{12\cdot 25}(133\alpha_1^2+82\alpha_2^2-283\alpha_1\alpha_2),\\
&E = \tfrac{1}{50}(-27\alpha_1^2-14\alpha_2^2+73\alpha_1\alpha_2).
\end{aligned}\tag{27}$$

Fourth order in u: We assume that the fourth-order term is equal to the gradient of

$$F\int_{-\infty}^{+\infty} u^3u_x^2\,dx.$$

This yields the following equations:

$$\begin{aligned}
&F = \tfrac{13}{50}\alpha_1^3+\tfrac{2}{25}\alpha_2^3-\tfrac{43}{150}\alpha_1^2\alpha_2-\tfrac{14}{75}\alpha_1\alpha_2^2,\\
&F = -\tfrac{7}{50}\alpha_1^3+\tfrac{4}{25}\alpha_2^3+\tfrac{71}{150}\alpha_1^2\alpha_2-\tfrac{47}{75}\alpha_1\alpha_2^2,\\
&0 = (\alpha_2-\tfrac{5}{2}\alpha_1)(\alpha_2-\alpha_1)(\alpha_2-2\alpha_1).
\end{aligned}\tag{28}$$

Combining the first and second equation for F one gets a cubic equation having the following solutions

$$\alpha_2 = \lambda\alpha_1,\ \lambda = 1,\ \tfrac{5}{2},\ 2.\tag{29}$$

Then the third equation is satisfied automatically. Now one can insert the three cases of (29) into (27) and (26) and obtain three sets of coefficients such that $J(u)K(u)$ is a gradient. Before we list these cases we would like to make a remark: The choice of our potentials seems to be rather special, but in fact one can show that all the possible potentials have to be of the form we assumed.

First case:

$$\alpha_2=\alpha_1;\ \alpha_3=\tfrac{1}{5}\alpha_1^2;\ \alpha=\tfrac{2}{5}\alpha_1;\ \beta=\tfrac{1}{25}\alpha_1^2;\ \delta=\gamma=0.$$

These coefficients lead to the operators (2) and (3) and Eq. (1) (putting $\alpha_1=\frac{5}{2}\zeta$).

Second case:

$$\alpha_2=\tfrac{5}{2}\alpha_1;\ \alpha_3=\tfrac{1}{5}\alpha_1^2;\ \alpha=\tfrac{2}{5}\alpha_1;\ \gamma=\tfrac{2}{5}\alpha_1;\ \beta=\tfrac{4}{25}\alpha_1^2;\ \delta=0.$$

This leads immediately to the case considered in Sec. I B (where $\alpha_1=\frac{5}{2}\zeta$).

Third case:

$$\alpha_2=2\alpha_1;\ \alpha_3=\tfrac{3}{10}\alpha_1^2;\ \alpha=\tfrac{2}{5}\alpha_1;\ \gamma=\tfrac{2}{5}\alpha_1;\ \beta=\tfrac{2}{25}\alpha_1^2;\ \delta=\tfrac{2}{25}\alpha_1^2.$$

This case leads to the first generalization of the Korteweg–de-Vries equation, a case which is not at all interesting since the bi-Hamiltonian structure of that equations is already known.

Having proved the assertions of Secs. IA and IB we obtain IC out of the general transformation formulas for bi-Hamiltonian systems presented in Ref. 7. The verification of the bi-Hamiltonian structure of Eq. (18) is, again, not at all difficult (once the coefficients are known). Another approach to Eq. (18) and its properties will be given in Sec. V.

V. HEREDITARINESS

Let us now turn our attention to the question whether or not the conserved covariants in (6b) and (19b) are closed, i.e., gradients of suitable potentials. Checking that directly leads into a rather frustrating adventure. But there is another method, namely, to determine whether or not the corresponding operator $\Phi(u)=\Theta(u)\,J(u)$ is hereditary. Before we explain what hereditariness can do for us, let us give the following abstract definition: The tensorfield given by Φ is said to be *hereditary* if, with respect to the Lie algebra of vector fields, the following holds[14]:

$$\Phi^2[H_1,H_2]+[\Phi H_1,\Phi H_2]=\Phi\{[H_1,\Phi H_2]+[\Phi H_1,H_2]\}$$

for all C^∞-vector fields H_1 and H_2 (compare with the notion of a Nijenhuis operator in Ref. 16, or a regular operator in Ref. 17).

Now, assume that (J, K, Θ) is bi-Hamiltonian and that K and $\Phi=\Theta J$ are invariant against x-translation. Invariance against x-translation means that

$$D\Phi(u)-\Phi(u)D=\frac{\partial}{\partial\epsilon}\Phi(u+\epsilon u_x)_{|\epsilon=0}\quad\text{for all } u\in S$$

and that the flow $u_t=K(u)$ commutes with x-translation. Under these assumptions the hereditariness of $\Phi=\Theta J$ has the following consequences (see Refs. 8, 13, or 14): The vector fields given by

$$\{\Phi(u)^n K(u)\,|\,n\in N_0\}\,U\,\{\Phi(u)^n u_x\,|\,n\in N_0\}\tag{30}$$

are contained in an abelian subalgebra of the Lie-algebra given by all vector fields. All the covector fields given by

$$\{J(u)\Phi^n(u)K(u)\,|\,n\in N_0\}\,U\,\{J(u)\Phi^n(u)u_x\,|\,n\in N_0\}\tag{31}$$

do have potentials, i.e., they are closed covector fields.

Now, all the operators Φ considered in Sec. I are in fact hereditary. To prove this is only necessary for one of the

operators, since all the corresponding bi-Hamiltonian systems are related by Bäcklund transformations (which preserve the property of hereditariness). To check the algebraic relation seems to be a rather straightforward calculation, but it is not, the calculation is cumbersome and extremely boring.

However, there are other ways to derive the hereditary property of Φ from. Let us indicate this briefly: For example, one can use the fact that any of the equations is an isospectral flow for some eigenvalue problem[5,6] and that the gradients of the corresponding eigenvalues are eigenvectors for $\Phi^{+}(u)$. From this one can show that Φ is hereditary.

Another procedure is given by the results of Ref. 21. There the authors showed that for certain dynamical systems the two given Hamiltonian structures are compatible. Since the CDGSK equation is the reduction of one of these systems one can derive that the two Hamiltonian structures of this equation are compatible. However, compatible Hamiltonian structures yield hereditary operators.[14]

Now, application of (30) immediately yields that the σ_{3n+1} and σ_{3n+3} (considered in Sec. I) are infinitesimal generators of an abelian symmetry group. Observation (31) yields that the corresponding G_{3n} and G_{3n+2} have potentials. Since J is a Lie-algebra homomorphism from the vector fields into the gradients (endowed with the Poisson brackets) all these conserved quantities must be in involution.

These results can also be obtained by other methods. The fact that the G have potentials can be derived by the ingenious reasoning presented in Ref. 20, and the abelian structure of the symmetry group can be derived from a general result of Tu.[22]

Finally, we would like to point out the connection with the soliton solutions mentioned in I D.

Let Φ be as above and consider the system

$$u_t = \Phi(u)u_x = K(u). \tag{32}$$

Then it can be shown[8,14] that the manifold

$$M = \left\{ u \in S \,|\, u_x = \sum_{k=1}^{N} \omega_k, \quad \omega_k \quad \text{eigenvector of } \Phi(u) \right\}$$

must be invariant under the flow given by (32). Let $u(t)$ be a pure N-soliton solution, i.e.,

$$u(x,t) = \sum_{n=1}^{N} u_n(x - \Theta_n - c_n t) + \Delta(x,t)$$

with $\|\Delta(\cdot,t)\| \to 0$ for $|t| \to \infty$ (in L^2-norm) and with

$$u_{n_x} = c_n K(u_n) \quad \text{for } n = 1,...,N. \tag{33}$$

Then, by (33), the u_{n_x} are eigenvectors of $\Phi(u_n)$ and, since $\Phi(u)$ is semilocal, $u(t)$ belongs, for $|t| = \infty$, (i.e., asymptotically) to the manifold M. Hence $u(t)$ belongs for all t to this manifold. By comparison of the dimensions one sees that M is exactly the manifold of N- soliton solutions [with prescribed asymptotic speeds $c_1,...,c_N$ depending on the eigenvalues of $\Phi(u)$].

ACKNOWLEDGMENTS

We are indebted to the referee for his critical and constructive remarks.

[1]P. J. Caudrey, R. K. Dodd, and J. D. Gibbon, Proc. R. Soc. London A **351**, 407 (1976).
[2]J. Satsuma and D. J. Kaup, J. Phys. Soc. Japan **43**, 692 (1977).
[3]R. K. Dodd and J. D. Gibbon, Proc. R. Soc. London A **358**, 287 (1977).
[4]K. Sawada and T. Kotera, Progr. Theor. Phys. **51**, 1355 (1974).
[5]A. P. Fordy and J. Gibbons, Phys. Lett. A **75**, 325 (1980).
[6]D. J. Kaup, Studies Appl. Math. **62**, 189 (1980).
[7]A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980).
[8]B. Fuchssteiner, Nonlinear Analysis **3**, 849 (1979).
[9]B. Fuchssteiner, Commun. Math. Phys. **55**, 187 (1977).
[10]J. Marsden, *Application of Global Analysis in Mathematical Physics* (Publish or Perish Inc., Boston, 1974).
[11]Y. Choquet-Bruhat, C. De Witte-Morette, and M. Dillard-Bleick: *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).
[12]B. Fuchssteiner, "The Lie algebra structure of generalized bi-Hamiltonian dynamical systems and their symmetry groups" (to appear).
[13]B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[14]B. Fuchssteiner, Prog. Theor. Phys. **65**, 861 (1981).
[15]P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[16]F. Magri, in *Lecture Notes in Physics*, edited by M. Boiti, F. Pimpinelli, and G. Soliani (Springer-Verlag, New York, 1980), Vol. 120, pp. 233–263.
[17]I. M. Gel'fand and I. Ya. Dorfman, Funct. Analy. i. Eqs. Priloz **13**(4), 13 (1979).
[18]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[19]A. S. Fokas and B. Fuchssteiner, Nonlinear Analysis **5**, 423 (1981).
[20]A. Chodos, Phys. Rev. D **21**, 2818 (1980).
[21]B. A. Kupershmidt and G. Wilson, Invent. Math. **62**, 403 (1981).
[22]G. Tu, Commun. Math. Phys. **77**, 289 (1980).

Singular symmetries of integrable curves and surfaces

B. A. Kupershmidt[a)]
Los Alamos Scientific Laboratory, CNLS, P. O. Box 1663, Los Alamos, New Mexico 87545[b)]

(Received 4 August 1981; accepted for publication 2 October 1981)

If $w = w(x,t)$ is a solution of $w_t = 6ww_x - w_{xxx} + 6\epsilon^2 w^2 w_x$ then $\overline{w} = -w - \epsilon^{-2}$ is also a solution. In general, integrable families and their members admit discrete symmetries whereas original systems may not.

PACS numbers: 02.30.Jr, 02.30.Qy

1. INTRODUCTION

As a rule, integrable systems (i.e., evolutions equations with an infinity of integrals) occur in families (see Refs. 1 and 2). The simplest example is provided by[3] the "Gardner transformation": If w is a solution of the Gardner equation

$$w_t = \partial(3w^2 - w_{xx} + 2\epsilon^2 w^3), \quad \partial \equiv d/dx \tag{1.1}$$

then u, produced by the map $g(\epsilon)$,

$$g(\epsilon): w \mapsto u = w + \epsilon^2 w^2 + \epsilon w_x, \tag{1.2}$$

satisfies the Korteweg–de Vries (KdV) equation

$$u_t = \partial(3u^2 - u_{xx}). \tag{1.3}$$

Now if w is a solution of (1.1) then $\overline{w}$,

$$\overline{w} = s(w) := -w - \epsilon^{-2}, \tag{1.4}$$

is also a solution of (1.1). Thus (1.1), which is a *regular* deformation of (1.3), admits the symmetry s which is *singular* in the deformation parameter ϵ. In particular, when ϵ tends to zero, s does not have any limit. This, together with the obvious absence of any (nontrivial) symmetries of the form $u \mapsto \overline{u}(u)$ for the KdV equation, indicates that the appearance of new symmetries, when an integrable system is included in an integrable family, might be a generic event. Below we prove this for deformations of higher KdV equations constructed in Ref. 1, and for some other systems considered in Ref. 2, such as deformation of the MKdV equation

$$v_t = 6v^2 v_x - v_{xxx}, \tag{1.5}$$

which *does* admit a (discrete) symmetry

$$m: v \mapsto \overline{v} = -v. \tag{1.6}$$

2. HIGHER KdV CURVES

The KdV equation (1.3) is just one of the series of equations resulting from the Lax representation

$$L_t = [P, L] \tag{2.1}$$

with $L = -\partial^2 + u$. Each of these equations is a point on a curve of integrable systems. This curve can be described in the following way, somewhat simpler than in Ref. 1.

If v is a solution of MKdV$_r$ (i.e., higher modified KdV equation # r) (then "Miura transformation")

$$M: v \mapsto u = v^2 + v_x \tag{2.2}$$

is a solution of KdV$_r$. More generally, if v is a solution of a linear combination of MKdV fields, then (2.2) is a solution of a corresponding combination of KdV fields. In addition, if u is a solution of linear combination of KdV fields, then $\tilde{u} = u + c$, $c = \text{const}$, is again a solution (of another combination) (see, e.g., Ref. 4). Thus for $c = -\epsilon^{-2}/4$ and

$$w = \epsilon^{-1} v - \epsilon^{-2}/2, \tag{2.3}$$

(2.2) becomes

$$\tilde{u} = w + \epsilon^2 w^2 + \epsilon w_x, \tag{2.4}$$

which is just (1.2) if we drop the tilde off $\tilde{u}$. Therefore, the w curve through the KdV$_r$ is a linear combination of MKdV fields combined with the affine transformation (2.3). This curve depends polynomially upon ϵ.[1] Let us write $w \in \text{KdV}_r(\epsilon)$ to mean that w is a solution of a deformed KdV$_r$ with parameter ϵ.

Theorem 2.5: (a) The map s from (1.4) is a symmetry of KdV$_r(\epsilon)$; (b) KdV$_r(\epsilon)$ depends on ϵ^2; (c)

$$g(-\epsilon) \circ s = g(\epsilon). \tag{2.6}$$

Proof: (a) If v is a solution of a linear combination of MKdV fields, then $\overline{v} = -v$ is also a solution (see, e.g., Ref. 5). So (2.3) gives $\overline{w} = (\epsilon^{-1}\overline{v} - \epsilon^{-2}/2) = (-\epsilon^{-1} v - \epsilon^{-2}/2)$ $= -(\epsilon^{-1} v - \epsilon^{-2}/2) - \epsilon^{-2} = -w - \epsilon^{-2}$; (c) $[g(-\epsilon) \circ s](w) = (g(-\epsilon))(-w - \epsilon^{-2})$ $= -w - \epsilon^{-2} + \epsilon^2(-w - \epsilon^{-2})^2 - \epsilon(-w - \epsilon^{-2})_x$ $= w + \epsilon^2 w^2 + \epsilon w_x = (g(\epsilon))(w)$; (b) since $\tilde{u} = u - \epsilon^{-2}/4$, we start from linear combinations of KdV and MKdV fields, with coefficients which are even in ϵ. Now (2.3) tells us that $v = \epsilon(w + \epsilon^{-2}/2)$, so that if $\dot{v} = F(v)$ then $\dot{w} = \epsilon^{-1}\dot{v} = \epsilon^{-1} F[\epsilon(w + \epsilon^{-2}/2)]$ and since F is an odd function, $\dot{w}$ is even in ϵ. ∎

Theorem 2.5(b) immediately provides Bäcklund transformations, that is, infinitesimal formal automorphisms, of all KdV$_r$ eqs. Indeed, KdV$_r(\epsilon)$ depends upon ϵ^2; therefore if $w \in \text{KdV}_r(\epsilon)$ then both $(g(\epsilon))(w)$ and $(g(-\epsilon))(w)$ are solutions of KdV$_r$, and we get $g(-\epsilon) \circ g(\epsilon)^{-1}$ as the desired automorphism (cf. tedious constructions in Ref. 2).

3. HIGHER MKdV CURVES

To understand the degree of generality of the appearance of new symmetries, it is natural to look outside the (very degenerate, from every point of view) KdV situation. In this section we analyze MKdV$_r$ fields.

To see the pattern let us begin from the simplest case: the MKdV equation[2]

$$v_t = 6v^2 v_x - v_{xxx}, \tag{3.1}$$

[a)]Supported in part by NSF Grant MCS 800 3104.

[b)]Permanent address: Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109.

 0022-2488/82/030364-03$02.50 © 1982 American Institute of Physics

its deformation

$$q_t = 6[\mathrm{sh}2\epsilon q/2\epsilon]^2 q_x - q_{xxx} + 2\epsilon^2 q_x^3, \tag{3.2}$$

its reduction $Mg(\epsilon)$

$$Mg(\epsilon): q \to v = \mathrm{sh}2\epsilon q/2\epsilon + \epsilon q_x, \tag{3.3}$$

and the deformation of the Miura map (2.2)

$$M(\epsilon): q \to w = (\mathrm{sh}\epsilon q/\epsilon)^2 + q_x. \tag{3.4}$$

This means: if q satisfies MKdV(ϵ) (3.2), then (3.3) satisfies MKdV (3.1), (3.4) satisfies KdV(ϵ) (1.1), and

$$M \circ Mg(\epsilon) = g(\epsilon) \circ M(\epsilon). \tag{3.5}$$

It is easy to compute symmetries of (3.2). They have the form

$$m_1(k): q \to q + \pi i k/2\epsilon,\ m_2(k): q \to -q + \pi i k/2\epsilon,\ k \in \mathbb{Z}. \tag{3.6}$$

Notice that $m_2(0)$ covers m of (1.6):

$$Mg(\epsilon) \circ m_2(0) = m \circ Mg(\epsilon), \tag{3.7}$$

so the involution m does survive the deformation (3.2).

I now claim that (3.6) provides symmetries not only of MKdV(ϵ) but also of $\mathrm{MKdV}_r(\epsilon)$ for all r's as well. First, we define MKdV_r simply as $[Mg(\epsilon)]^{-1}(\mathrm{MKdV}_r)$.

Proposition 3.8: $\mathrm{MKdV}_r(\epsilon)$ depends on ϵ^2.

Proof: By (3.5), $\mathrm{MKdV}_r(\epsilon) = [M(\epsilon)]^{-1}(\mathrm{KdV}_r(\epsilon))$. But $\mathrm{KdV}_r(\epsilon)$ depends upon ϵ^2 by Theorem 2.5(b), and $M(-\epsilon) = M(\epsilon)$ by (3.4). ∎

Corollary 3.9: $Mg(\epsilon) \circ [Mg(-\epsilon)]^{-1}$ yields the Bäcklund transformation for all MKdV_r fields.

Lemma 3.10: (a) $(Mg[(-1)^{j+1}\epsilon]) \circ m_j(k) = Mg(\epsilon)$ for $j + k \equiv 1 \pmod 2$; (b) $(Mg[(-1)^j \epsilon]) \circ m_j(k) = m \circ Mg(\epsilon)$ for $j + k \equiv 0 \pmod 2$.

Proof: $((Mg[(-1)^s\epsilon]) \circ m_j(k))(q)$
$= (Mg[(-1)^s\epsilon])[(-1)^{j+1}q + \pi i k/2\epsilon]$
$= (-1)^{k+j+1}\mathrm{sh}2\epsilon q/2\epsilon + (-1)^{s+j+1}\epsilon q_x$. ∎

Theorem 3.11: $\mathrm{MKdV}_r(\epsilon)$ possesses symmetry (3.6).

Proof: Let $q \in \mathrm{MKdV}_r(\epsilon)$. If $j + k \equiv 1 \pmod 2$ then $m_j(k)(q) = ((Mg[(-1)^{j+1}(\epsilon)])^{-1} \circ Mg(\epsilon))(q) \in \mathrm{MKdV}_r(\epsilon)$ by Proposition (3.8). Analogously for $j + k \equiv 0 \pmod 2$. ∎

4. RELATIONS BETWEEN SYMMETRIES OF KdV AND MKdV CURVES

Symmetries (1.4) and (3.6) are closely connected, as one would expect.

Theorem 4.1: For $k \equiv 1 \pmod 2$, $M(\epsilon) \circ m_2(k) = s \circ M(\epsilon)$.

Proof: $(s \circ M(\epsilon))(q) = s(\mathrm{sh}^2\epsilon q/\epsilon^2 + q_x)$
$= -\mathrm{sh}^2\epsilon q/\epsilon^2 - q_x - \epsilon^{-2}$
$= -\mathrm{ch}2\epsilon q/2\epsilon^2 - q_x - \epsilon^{-2}/2, (M(\epsilon) \circ m_2(k))(q)$
$= (M(\epsilon))(-q + \pi i k/2\epsilon)$
$= [\mathrm{ch}2\epsilon(-q + \pi i k/2\epsilon) - 1]/2\epsilon^2 + (-q + \pi i k/2\epsilon)_x$
$= (-1)^k \mathrm{ch}2\epsilon q/2\epsilon^2 - 1/2\epsilon^2 - q_x$. ∎

5. THE KdV SURFACE—FINITE CASE

Let us see what happens when one has a "deformation of a deformation", i.e., a deformation depending upon two parameters. In this section we consider the case in which the deformation is finite, that is, involves only a bounded number of derivatives.

Our system now is $\mathrm{KdV}(\epsilon,\nu)$,[1]

$$\begin{cases} p_t = 6p_x(1 + \epsilon^2 C) - p_{xxx} + 2\omega^2 p_x^3, & \omega = \epsilon\nu, \\ C = C(\epsilon,\nu,p) = \mathrm{sh}2\omega p/2\omega + (\mathrm{ch}2\omega p - 1)/2\epsilon^2, \end{cases} \tag{5.1}$$

together with its reduction

$$G(\epsilon,\nu): p \to w = C(\epsilon,\omega,p) + \nu p_x, \tag{5.2}$$

meaning: $p \in \mathrm{KdV}(\epsilon,\nu) \Rightarrow w = G(\epsilon,\nu)(p)$ satisfies (1.1). Notice that $G(0,\nu) = g(\nu)$, thus (5.2) represents a deformation of the reduction (1.2).

Let us now look for symmetries of (5.1). Clearly, $\bar{p} = f(p)$ is again a solution of (5.1) iff

$$f(p) = \alpha p + \beta, \quad \alpha = \pm 1, \quad \beta = \mathrm{const},$$
$$C(1 + \epsilon^2 C) = \bar{C}(1 + \epsilon^2 \bar{C}). \tag{5.3}$$

Therefore we get four cases: $\alpha = \pm 1$, $\bar{C} = C$, or $\bar{C} = -C - \epsilon^{-2}$. Working them out one easily gets symmetries

$$M_1(k): p \to \bar{p} = p + \pi i k/2\epsilon\nu,\ k \in \mathbb{Z} \quad (\alpha = 1); \tag{5.4}$$

$$\begin{cases} M_3: p \to \bar{p} = -p + \beta_3, & \mathrm{sh}2\epsilon\nu\beta_3 = 2\epsilon\nu/(\epsilon^2 - \nu^2), \\ \mathrm{ch}2\epsilon\nu\beta_3 = (\nu^2 + \epsilon^2)/(\nu^2 - \epsilon^2) & (\bar{C} = C), \end{cases} \tag{5.5}$$

$$\begin{cases} M_4: p \to \bar{p} = -p + \beta_4, & \mathrm{sh}2\epsilon\nu\beta_4 = 2\epsilon\nu/(\nu^2 - \epsilon^2), \\ \mathrm{ch}2\epsilon\nu\beta_4 = (\nu^2 + \epsilon^2)/(\epsilon^2 - \nu^2) & (\bar{C} = -C - \epsilon^{-2}). \end{cases} \tag{5.6}$$

Notice that the β's in (5.5) and (5.6) are defined modulo $i\pi\epsilon^{-1}\nu^{-1}\mathbb{Z}$, and whereas (5.4) is rational in the deformation parameters, as were (1.4) and (3.6), (5.5) and (5.6) are no longer rational (or even algebraic). Notice also that for $\nu \to 0$, one branch of β_4 tends to $-\epsilon^{-2}$ which results in that branch of (5.6) being a *regular* (in ν) deformation of the singular (in ϵ) symmetry s in (1.4). [The same applies to M_3 when $\epsilon \to 0$ since $\beta_3(\epsilon,\nu) = \beta_4(\nu,\epsilon)$.] This observation reflects general relations between symmetries of the $\mathrm{KdV}(\epsilon,\nu)$ and $\mathrm{KdV}(\epsilon)$

Proposition 5.7: (a) $G(\epsilon,\nu) \circ M_1(k) = G(\epsilon,\nu)$ for $k \equiv 0 \pmod 2$; (b) $G(\epsilon,-\nu) \circ M_1(k) = s \circ G(\epsilon,\nu)$ for $k \equiv 1 \pmod 2$; (c) $G(\epsilon,-\nu) \circ M_3 = G(\epsilon,\nu)$; (d) $G(\epsilon,\nu) \circ M_4 = s \circ G(\epsilon,\nu)$.

Proof: (a) and (b) $(M_1(k))(p) = p + \pi i k/2\epsilon\nu$, so
$(G(\epsilon, \pm\nu) \circ M_1(k))(p) = (-1)^k \mathrm{sh}2\epsilon\nu p/2\epsilon\nu$
$+ [(-1)^k \mathrm{ch}2\epsilon\nu p - 1]/2\epsilon^2 \pm \nu p_x$,
(c) $(G(\epsilon,-\nu) \circ M_3)(p) = (G(\epsilon,-\nu))(-p + \beta_3)$
$= \nu p_x + \bar{C} = (G(\epsilon,\nu))(p)$ by (5.5), (d)
$(G(\epsilon,\nu) \circ M_4)(p) = (G(\epsilon,\nu))(-p + \beta_4)$
$= -\nu p_x + \bar{C} = -\nu p_x - C - \epsilon^{-2} = -(\nu p_x + C) - \epsilon^{-2}$
$= s \circ G(\epsilon,\nu))(p)$ by (5.6). ∎

6. THE KdV SURFACE—INFINITE CASE

Let us consider now a situation when an undeformed equation itself contains an infinite number of derivatives, the finite depth (FD) equation[6]:

$$U_t = 2UU_x + T(U_{xx}), \tag{6.1}$$

where $T \in \mathbb{Q}[[\delta^2, \partial]]$ is defined by

$$T = \delta^{-1} \frac{e^{\delta\partial} + e^{-\delta\partial}}{e^{\delta\partial} - e^{-\delta\partial}} - \delta^{-2}\partial^{-1} = \tfrac{1}{3}\partial + O(\delta^2). \tag{6.2}$$

Thus for $\delta = 0$ we end up with the KdV equation (with inessential rescaling of variables). The FD equation has the

deformation

$$Q_t = Q_x[2\delta\epsilon Q - (1-\delta/\epsilon)(-1+\exp 2\delta\epsilon Q)]/\delta^2 + TQ_{xx} + \epsilon\delta Q_x TQ_x, \tag{6.3}$$

and the reduction

$$R(\epsilon,\delta): Q \to U = [2\delta\epsilon Q - (1-\delta/\epsilon)(-1+\exp 2\delta\epsilon Q)]/2\delta^2 - \epsilon Q_x + \epsilon\delta TQ_x, \tag{6.4}$$

which are reduced, for $\delta = 0$ to

$$Q_t = 2Q_x(Q - \epsilon^2 Q^2) + Q_{xxx}/3, \tag{6.3'}$$

$$R(\epsilon,0): Q \to U = Q - \epsilon Q_x - \epsilon^2 Q^2, \tag{6.4'}$$

which are exactly (1.1) and (1.2) (again with a rescaling).

It is easy to find all symmetries of (6.3); as we shall see below, if there are no relations between the parameters ϵ and δ, then there are no symmetries (except the identity, naturally). Thus one is forced to extend the notion of symmetry to allow parameters (on a curve or surface) to *change after symmetries*.

In our situation (6.3), we thus look for maps $A: Q \to \bar{Q} = f(Q)$ such that if Q satisfies (6.3) with parameters (ϵ,δ) then $\bar{Q}$ satisfies (6.3) with parameters $(\bar{\epsilon},\bar{\delta})$. Performing standard computations we obtain

$$f(Q) = \alpha Q + \beta, \quad \bar{\delta}^2 = \delta^2, \quad \alpha\bar{\epsilon}\bar{\delta} = \epsilon\delta, \quad \beta = \alpha(\alpha-1)/2\epsilon^2, \tag{6.5}$$

$$(1-\delta/\epsilon)\exp(\delta/\epsilon) = (1-\alpha\delta/\epsilon)\exp(\alpha\delta/\epsilon). \tag{6.6}$$

Thus either $\alpha = 1, \beta = 0$ (identical transformation), or α is regular in δ/ϵ and

$$\alpha = -1 + O(\delta/\epsilon), \tag{6.7}$$

as guaranteed by the implicit function theorem applied to (6.6). In this case (6.5) yields $\beta = \epsilon^{-2}[1 + O(\delta/\epsilon)]$, so we get

$$A: Q \to \bar{Q} = [-1 + O(\delta/\epsilon)]Q + \epsilon^{-2}[1 + O(\delta/\epsilon)], \tag{6.8}$$

which thus is a deformation of the symmetry $Q \to \bar{Q} = -Q + \epsilon^{-2}$ of (6.3') *regular* in δ. Notice that $A \circ A = \mathrm{id}$ as well as $s \circ s$. Finally, observe that there is also a deformational analog of (2.6):

$$R(\bar{\epsilon},\bar{\delta}) \circ A = R(\epsilon\delta/\bar{\delta}, \delta\delta/\bar{\delta}) \tag{6.9}$$

(recall that $\delta/\bar{\delta} = \pm 1$).

7. DISCUSSION

Curves and surfaces which deform a given integrable system acquire new symmetries which may depend upon deformation parameters in regular, singular, rational, and irrational manners. In addition, symmetries can be accompanied by a reparametrization. As a rule, these symmetries are compatible with morphisms and deformations of morphisms between different systems.

When the operator $L = \partial^{n+2} + \sum_{i=0}^{n} u_i \partial^i$ in (2.1) has order $\geqslant 3$, the situation becomes less transparent. Deformations and Bäcklund transformations do exist,[2] but not much is known about them. Here I mention the first nontrivial case of $L = \partial^3 + \cdots$, $P = \partial^2 + \cdots$, the so-called Boussinesq equation. Its deformation (Chap. III, Ref. 2) has the form

$$\begin{cases} \gamma_t = \partial(\gamma_x - 2\delta), \\ \delta_t = \partial(\gamma^2/3 + 2\gamma_{xx}/3 - \delta_x) - \epsilon^2\gamma^2\gamma_x + \epsilon\gamma_x(2\delta - \gamma_x), \end{cases} \tag{7.1}$$

and it is simple to verify that

$$B: \gamma \to \bar{\gamma} = -\gamma + 2\epsilon^{-2}/3, \quad \delta \to \bar{\delta} = -\delta \tag{7.2}$$

is indeed the symmetry (and involution) of (7.1) with *reparametrization* $\epsilon \to -\epsilon$.

At last, let me briefly comment on the ODE situation. Let us take as the show-room study the usual Toda lattice (TL)

$$u_t = \alpha(v - v^-), \quad v_t = \alpha v(u^+ - u), \quad \alpha = \pm 1, \tag{7.3}$$

where equalities in (7.3) are understood as functions on $\mathbf{Z}$ (or $\mathbf{Z}/N\mathbf{Z}$) and can be read as $u_{n,t} = \alpha(v_n - v_{n-1})$, $v_{n,t} = \alpha v_n(u_{n+1} - u_n)$. Since TL (and the whole hierarchy possesses three Hamiltonian structures,[7] it is easy to write down corresponding TL curves and surfaces,

$$p_t = \alpha(1+\epsilon p)(q - q^-), \quad q_t = \alpha q(p^+ - p), \quad \alpha = \pm 1 \tag{7.4}$$

$$b_1: u = p + \epsilon q, \quad v = q + \epsilon p^+ q, \tag{7.5}$$

$$b_2: u = p + \epsilon q^-, \quad v = q + \epsilon pq, \tag{7.6}$$

$$\begin{cases} P_t = \alpha(1+\nu P)(1+\epsilon P)(Q - Q^-), \\ Q_t = \alpha Q(1+\epsilon\nu Q)(P^+ - P), \end{cases} \tag{7.7}$$

$$B_1: p = P + \nu Q(1+\epsilon P), \quad q = Q + \nu P^+ Q,$$

$$B_2: p = P + \nu Q^-(1+\epsilon P), \quad q = Q + \nu PQ. \tag{7.8}$$

Notice the difference from the KdV case: TL(ϵ) (7.4) is a bi-Hamiltonian and TL(ϵ,ν) (7.7) is a Hamiltonian system, whereas KdV (ϵ) has one Hamiltonian structure and KdV(ϵ,ν) has none.

Typical symmetries are: for TL(ϵ)

$$\bar{p} = -p - 2/\epsilon, \quad \bar{q} = -q, \quad \bar{\alpha} = -\alpha, \tag{7.9}$$

and for TL(ϵ,ν)

$$\bar{P} = P, \quad \bar{Q} = -Q - 1/\epsilon\nu, \quad \bar{\alpha} = -\alpha. \tag{7.10}$$

We see that α changes sign, i.e., we have to change t to $-t$.

ACKNOWLEDGMENTS

The author appreciates valuable comments made by D. Burns, I. Dolgachev, Yu. I. Manin, and A. Greenspoon.

[1] B. A. Kupershmidt, J. Math. Phys. **22**, 449 (1981).
[2] B. A. Kupershmidt, "Deformations of Integrable Systems," Proc. R. Irish Acad. (to appear).
[3] R. M. Miura, SIAM Rev. **18**, 412 (1976).
[4] B. A. Kupershmidt and G. Wilson, Invent. Math. **62**, 403 (1981).
[5] M. Adler and J. Moser, Commun. Math. Phys. **61**, 1 (1978).
[6] J. Gibbons and B. Kupershmidt, Phys. Lett. A **79**, 31 (1980).
[7] B. A. Kupershmidt, "Mechanical systems with infinite number of particles and their continuous limits" (to appear).

Recurrence relations for the coefficients of perturbation expansions

H. J. W. Müller-Kirsten and L. K. Sharma [a)]
Department of Physics, University of Kaiserslautern, 6750 Kaiserslautern, West Germany

(Received 17 August 1981; accepted for publication 18 September 1981)

Exact recurrence relations are derived for the coefficients of the perturbation expansion of the Schrödinger wavefunction for large classes of potentials. The terms of the eigenvalue expansion can then be expressed in terms of these coefficients which therefore allow other investigations such as the large-order behavior of the expansion.

PACS numbers: 02.30.Lt, 02.30.Mu

1. INTRODUCTION

It was shown by Dingle[1] that before an asymptotic expansion can be used to yield exact or almost exact results, it is essential to know the behavior of its late terms since this determines the Borel summability of the expansion. The derivation of this large-order behavior is not a simple task. In quantum mechanical problems the most straightforward way to derive this behavior is by finding an approximate solution to the recurrence relation of the coefficients of the perturbation expansion of the appropriate wavefunction. This is the procedure which has been adopted in Refs. 2–5. This is therefore one reason why these recurrence relations are of considerable interest. Another reason for studying the recurrence relations of perturbation coefficients is that they can be handled by a computer relatively easily and hence can be used for purely numerical calculations.

The usual type of perturbation theory as expounded, e.g., in texts on quantum mechanics, does not lend itself readily to the derivation of the recurrence relation of its perturbation coefficients. The reason is that this type of perturbation theory, generally aimed at the calculation of a second- or at most third-order contribution, does not exploit the systematics inherent in a perturbation expansion as a result of the recurrence relation of its unperturbed contribution. It is only if this systematics is taken into account that a recurrence relation of the full perturbation coefficients can be written down. A perturbation procedure which makes ample use of this systematics has been described in numerous publications.[2–5] In these publications, of course, the recurrence relation of the perturbation coefficients (if at all mentioned) has simply been written down as a result of considerable familiarity with the perturbation procedure itself. Since the formulation of these recurrence relations is by no means trivial we derive them here for three cases which cover a large variety of problems. For each of these cases the derivation of the systematic perturbation expansion is available in the literature[2,4,5] and thus need not be reproduced here. The cases we consider are (a) a generalized anharmonic potential, (b) a generalized Yukawa potential, and (c) a periodic potential. It will be shown here that these recurrence relations are all of the same general type and that the resulting characteristic equation determining the eigenvalues can be expressed in terms of their solutions. Finally in Sec. 5 we make some concluding remarks.

For completeness we mention that the solution of recurrence relations has recently also been investigated by other authors[6,7] in order to obtain complete solutions of the Schrödinger equation. These authors, however, are primarily concerned with power series expansions. In order to relate these to our type of expansions discussed here it is necessary to sum a selected number of terms from each iteration.

2. THE GENERALIZED ANHARMONIC POTENTIAL

In Refs. 3 and 4 the solution $\psi = r^{1/2}\phi$ of the radial Schrödinger equation was investigated for potentials $V(r)$ which can be expanded in the neighborhood of a minimum. In particular it was shown that if one sets $r = e^z$ $(-\infty < Z < \infty)$ and writes

$$v(Z) \equiv \left[\frac{2\mu}{\hbar^2} r^2 (E - V(r))\right]_{r=e^z} = v(Z_0) + \sum_{i=2}^{\infty} \frac{(Z - Z_0)^i v^{(i)}(Z_0)}{i!},$$

where $Z = Z_0$ is the position of the minimum, the Schrödinger equation can be cast into a new form involving an (approximately) odd integer q and the redefined eigenvalue Δ, i.e.,

$$\mathscr{D}_q \phi = 2\Delta\phi - \sum_{i=3}^{\infty} \frac{v^{(i)}(Z_0)}{v^{(2)}(Z_0)} \frac{\omega^i}{i! h^{i-2}} \phi. \tag{1}$$

Here[3]

$$h = [-2v^{(2)}(Z_0)]^{1/4},$$
$$\omega = h(Z - Z_0),$$
$$\mathscr{D}_q \equiv -2d^2/d\omega^2 + 1/2\omega^2 - q,$$
$$\text{and } \Delta = (1/h^2)[v(Z_0) - (l + \tfrac{1}{2})^2] - 1/2q. \tag{2}$$

It was also shown in Ref. 3 that in a small domain around the minimum the solution ϕ can be expressed in terms of parabolic cylinder functions $\phi_q(Z) = D_{(q-1)/2}(\omega)$, i.e.,

$$\phi(Z) = \phi_q(Z) + \sum_{i=3}^{\infty} \sum_{\substack{j=2i,2i-4,\cdots \\ j\neq 0}}^{-2i} P_{i-2}(q,j)\phi_{q+j}(Z), \tag{3}$$

where (for convenience) a factor $1/h^{i-2}$ has been absorbed in the coefficients P_{i-2}. The first few of these coefficients can be read off the expansion given in Ref. 3. Our objective here is to derive the recurrence relation of the coefficients P_i.

[a)] On leave from Department of Applied Physics, Government Engineering College, Jabalpur, India.

 0022-2488/82/030367-06$02.50 © 1982 American Institute of Physics

Applying $\mathscr{D}_q$ to ϕ and using $\mathscr{D}_q\phi_q = 0$ and $\mathscr{D}_q\phi_{q+j} = j\phi_{q+j}$, one obtains

$$\sum_{i=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i} jP_{i-2}(q,j)\phi_{q+j}$$

$$= 2\Delta\left[\phi_q + \sum_{i=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i} P_{i-2}(q,j)\phi_{q+j}\right] - \sum_{i=3}^{\infty}\frac{v^{(i)}(Z_0)}{v^{(2)}(Z_0)}\frac{1}{i!h^{i-2}}\sum_{j=2i,2i-4,\cdots}^{-2i} S_i(q,j)\phi_{q+j}$$

$$- \sum_{i=3}^{\infty}\frac{v^{(i)}(Z_0)}{v^{(2)}(Z_0)}\frac{1}{i!h^{i-2}}\sum_{i'=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i} P_{i'-2}(q,j)\sum_{j'=2i,\cdots}^{-2i} S_i(q+j,j')\phi_{q+j+j'}, \tag{4}$$

where we have used the recurrence relation for parabolic cylinder functions, i.e.,

$$\omega^i\phi_q = \sum_{j=2i,2i-4,\cdots}^{-2i} S_i(q,j)\phi_{q+j},$$

where the coefficients S_i are easily calculated. With the definition

$$i\geqslant 3,\ -2i<j<2i,$$

$$[q,q+j]_{i-2} = -\frac{v^{(i)}(Z_0)}{v^{(2)}(Z_0)}\frac{1}{i!}S_i(q,j), \tag{5a}$$

except for $i=3, j=0$, in which case

$$[q,q]_1 = 2\Delta h - \frac{v^{(3)}(Z_0)}{v^{(2)}(Z_0)}\frac{1}{3!}S_3(q,j), \tag{5b}$$

the relation (4) can be written

$$\sum_{i=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i} jP_{i-2}(q,j)\phi_{q+j}$$

$$= \left\{\sum_{i=3}^{\infty}\frac{[q,q]_{i-2}}{h^{i-2}} + \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=3}^{\infty}\sum_{j=2i',2i'-4,\cdots}^{-2i'} P_{i'-2}(q,j)[q+j,q]_{i-2}\right\}\phi_q$$

$$+ \sum_{i=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i}\frac{[q,q+j]_{i-2}}{h^{i-2}}\phi_{q+j} + \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=3}^{\infty}\sum_{\substack{j=2i',\cdots\\ j\neq 0}}^{-2i'}\sum_{\substack{j'=2i,\cdots\\ j+j'\neq 0}}^{-2i} P_{i'-2}(q,j)[q+j,q+j+j']_{i-2}\phi_{q+j+j'}. \tag{6}$$

The term $2\Delta\phi_q$ is contained in $[q,q]_1/h$, and the term $2\Delta\sum_{i'}\sum_{j'}P_{i'-2}(q,j')\phi_{q+j'}$ in the last sum for $i=3, j'=0$. If we included in the last sum the term for $j=0$ we would have the additional contribution

$$\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=3}^{\infty}\sum_{\substack{j'=2i,\cdots\\ j'\neq 0}}^{-2i} P_{i'-2}(q,0)[q,q+j']_{i-2}\phi_{q+j'}. \tag{7}$$

With $P_{i'-2}(q,0) = \delta_{i'-2,0}$ this is zero.

However, including the term for $i'=2$ means including

$$\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{\substack{j'=2i,\cdots\\ j'\neq 0}}^{-2i}[q,q+j']_{i-2}\phi_{q+j'},$$

which is precisely the term preceding our sum in (6). Thus, we can write

$$\sum_{i=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i} jP_{i-2}(q,j)\phi_{q+j}$$

$$= \{\quad\}\phi_q + \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=2}^{\infty}\sum_{j=2i',\cdots}^{-2i'}\sum_{\substack{j'=2i,\cdots\\ j+j'\neq\phi}}^{-2i} P_{i'-2}(q,j)[q+j,q+j+j']_{i-2}\phi_{q+j+j'}. \tag{8}$$

We now demand that $\{\quad\}$ be zero, yielding an equation from which Δ is to be determined—thus

$$0 = \sum_{i=3}^{\infty}\frac{[q,q]_{i-2}}{h^{i-2}} + \sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=3}^{\infty}\sum_{j=2i',2i'-4,\cdots}^{-2i'} P_{i'-2}(q,j)[q+j,q]_{i-2}. \tag{9}$$

Equation (8) can then be written

$$\sum_{i=3}^{\infty}\sum_{\substack{j=2i,2i-4,\cdots\\ j\neq 0}}^{-2i} jP_{i-2}(q,j)\phi_{q+j}$$

$$=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=2}^{\infty}\sum_{j=2i'}^{-2i'}\sum_{\substack{j'-j=2i\\ j'\neq 0}}^{-2i'} P_{i'-2}(q,j)[q+j,q+j']_{i-2}\phi_{q+j'}$$

$$=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=2}^{\infty}\sum_{j=2i'}^{-2i'}\sum_{\substack{j'=2i+2i'\\ j'\neq 0}}^{-2i-2i'} P_{i'-2}(q,j)[q+j,q+j']_{i-2}\phi_{q+j'}$$

$$=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{i'=i}^{\infty}\sum_{j=2(i'-i+2)}^{-2(i'-i+2)}\sum_{\substack{j'=2(i'+2)\\ j'\neq 0}}^{-2(i'+2)} P_{i'-i}(q,j)[q+j,q+j']_{i-2}\phi_{q+j'}$$

$$=\sum_{i'=3}^{\infty}\sum_{\substack{j'=2(i'+2)\\ j'\neq 0}}^{-2(i'+2)}\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2(i'-i+2)}^{-2(i'-i+2)} P_{i'-i}(q,j)[q+j,q+j']_{i-2}\phi_{q+j'}$$

$$=\sum_{i'=3}^{\infty}\sum_{\substack{j'=2(i'+2)\\ j'\neq 0}}^{-2(i'+2)}\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j+j'=2(i'-i+2)}^{-2(i'-i+2)} P_{i'-i}(q,j+j')[q+j+j',q+j']_{i-2}\phi_{q+j'}$$

$$=\sum_{i'=3}^{\infty}\sum_{\substack{j'=2i'\\ j'\neq 0}}^{-2i'}\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=+2i,2i-4,\cdots}^{-2i} P_{i'-i}(q,j+j')[q+j+j',q+j']_{i-2}\phi_{q+j'}$$

since $P_{i-2}(q,\pm 2i\pm 4)=0$.

The first two summations are now the same on both sides of the equation. We can therefore equate the coefficients of ϕ_{q+j} on both sides. Hence

$$tP_{r-2}(q,t)=\sum_{i=3}^{\infty}\frac{1}{h^{i-2}}\sum_{j=2i,2i-4,\cdots}^{-2i} P_{r-i}(q,j+t)[q+j+t,q+t]_{i-2}. \tag{10}$$

This is the relation given in Ref. 4. There, of course, it was written down from a consideration of all possible moves from q to $q+j$ in $i-2$ steps.

3. THE GENERALIZED YUKAWA POTENTIAL

In Ref. 5 high-energy asymptotic solutions Ψ were derived for the radial Schrödinger equation containing a generalized Yukawa potential of the form

$$V(r)=\sum_{i=-1}^{\infty} M_{i+1}(-r)^i, \tag{11}$$

where all M_i are real. In the equation it was found convenient to set

$$l+n+1=-\Delta(K)/2K,$$

so as to incorporate the asymptotic behavior of the Coulomb Regge poles. Here $E=k^2$, $K=ik$, $n=0,1,2,\cdots$ and Δ is an expansion in descending powers of K.

In order to cast the radial Schrödinger equation into a more tractable form one sets $Z=-2Kr$ and $\Psi(l,k;z)=ce^{-z/2}Z^{l+1}\chi(l,k;Z)$, c being a normalization constant. Then the Schrödinger equation for the potential (11) becomes

$$\mathscr{D}_a\chi=\frac{1}{2K}(M_0-\Delta(K))\chi+\frac{1}{2K}\sum_{i=1}^{\infty}(Z/2K)^iM_i\chi, \tag{12}$$

and the solution χ,

$$\chi(a,b;Z)=\psi(a,b;Z)+\sum_{i=2}^{\infty}\frac{1}{(2K)^i}\times\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)} P_i(a,j)\psi(a+j,b;Z), \tag{13}$$

where

$$\mathscr{D}_a=Z\frac{d^2}{dZ^2}+(b-z)\frac{d}{dZ}-a,$$

$$a=l+1+\Delta(K)/2K=-n, \tag{14}$$

and

$$b=2l+2=-2n-\Delta(K)/K.$$

The functions $\psi(a,b;Z)$ are truncated hypergeometric functions defined by $\mathscr{D}_a\psi(a,b;Z)=0$. Hence,

$$\chi^{(0)}=\psi(a,b;Z)\equiv\psi(a),$$

is the zeroth-order contribution of χ. Again we observe that $\mathscr{D}_a\psi(a+j)=j\psi(a+j)$. Inserting $\chi(a,b;Z)$ into (12) we obtain

$$\sum_{i=2}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)} jP_i(a,j)\psi(a+j)=\frac{(M_0-\Delta)}{2K}\left\{\psi(a)+\sum_{i=2}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)}P_i(a,j)\psi(a+j)\right\}$$
$$+\frac{1}{2K}\sum_{i=1}^{\infty}\frac{M_i}{(2K)^i}\sum_{j=-i}^{i}S_i(a,j)\psi(a+j)+\frac{1}{2K}\sum_{i=1}^{\infty}\frac{M_i}{(2K)^i}\sum_{i'=2}^{\infty}\frac{1}{(2K)^{i'}}\sum_{\substack{j=-(i'-1)\\ j\neq 0}}^{(i'-1)}P_{i'}(a,j)\sum_{j'=-i}^{i}S_i(a+j,j')\psi(a+j+j'), \quad (15)$$

where we have used the relation

$$Z^i\psi(a)=\sum_{j=-i}^{i}S_i(a,j)\psi(a+j) \quad (16)$$

which follows readily from the recurrence relation for confluent hypergeometric functions [hence the coefficients $S_i(a,j)$ are known].

Defining

$$[a,a]_1=M_0-\Delta(K),$$

and for $i\geqslant 1$,

$$[a,a+j]_{i+1}=M_iS_i(a,j)\quad(0\leqslant j\leqslant i), \quad (17)$$

Eq. (15) can be written

$$\sum_{i=2}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)}jP_i(a,j)\psi(a+j)$$
$$=\left\{\frac{1}{2K}\sum_{i=0}^{\infty}\frac{1}{(2K)^i}[a,a]_{i+1}+\frac{1}{2K}\sum_{i=1}^{\infty}\frac{1}{(2K)^i}\sum_{i'=2}^{\infty}\frac{1}{(2K)^{i'}}\sum_{\substack{j=-(i'-1)\\ j\neq 0}}^{(i'-1)}P_{i'}(a,j)[a+j,a]_{i+1}\right\}\psi(a)$$
$$+\frac{1}{2K}\sum_{i=1}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i}[a,a+j]_{i+1}\psi(a+j)$$
$$+\frac{1}{2K}\sum_{i=0}^{\infty}\frac{1}{(2K)^i}\sum_{i'=2}^{\infty}\frac{1}{(2K)^{i'}}\sum_{\substack{j=-(i'-1)\\ j\neq 0}}^{(i'-1)}\sum_{\substack{j'=-i\\ j+j'\neq 0}}^{i}P_{i'}(a,j)[a+j,a+j+j']_{i+1}\psi(a+j+j'). \quad (18)$$

The term $(M_0-\Delta)/(2K)\psi(a)$ is contained in the first sum for $i=0$, while the term

$$\frac{[a,a]_1}{2K}\sum_{i=2}^{\infty}\cdot\frac{1}{(2K)^i}\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)}P_i(a,j)\psi(a+j)$$

is included in the last sum for $i=0$ and $j'=0$. Now in the last sum on the right-hand side of Eq. (18), the condition $j\neq 0$ can be removed since for $i'=2,3,\cdots$, the coefficients $P_{i'}(a,0)$ are all zero. Similarly, including the term for $i'=1$, we obtain simply zero. However, including the term for $i'=0$ means including

$$\frac{1}{2K}\sum_{i=0}^{\infty}\frac{1}{(2K)^i}\sum_{j=1}^{-1}\sum_{\substack{j'=-i\\ j+j'\neq 0}}^{+i}P_0(a,j)[a+j,a+j+j']_{i+1}\psi(a+j+j'). \quad (19)$$

Since $P_0(a,j)=0$ except for $j=0$, expression (19) is

$$\frac{1}{2K}\sum_{i=0}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i}[a,a+j]_{i+1}\psi(a+j).$$

Since $[a,a+j]_1=0$ for $j\neq 0$, this is the same as the term preceding the last sum in Eq. (18). Thus we can write

$$\sum_{i=2}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)}jP_i(a,j)\psi(a+j)$$
$$=\{\quad\}\psi(a)+\frac{1}{2K}\sum_{i=0}^{\infty}\frac{1}{(2K)^i}\sum_{i'=0}^{\infty}\frac{1}{(2K)^{i'}}\sum_{j=-(i'-1)}^{(i'-1)}\sum_{\substack{j'=-i\\ j+j'\neq 0}}^{+i}P_{i'}(a,j)[a+j,a+j+j']_{i+1}\psi(a+j+j'). \quad (20)$$

As in the first example we set $\{\quad\}=0$ yielding an equation from which Δ and hence the eigenvalues are determined. Thus

$$0=\sum_{i=0}^{\infty}\frac{1}{(2K)^i}[a,a]_{i+1}+\sum_{i=1}^{\infty}\frac{1}{(2K)^i}\sum_{i'=2}^{\infty}\frac{1}{(2K)^{i'}}\sum_{\substack{j=-(i'-1)\\ j\neq 0}}^{(i'-1)}P_{i'}(a,j)[a+j,a]_{i+1}. \quad (21)$$

Equation (20) can now be written

$$\sum_{i=2}^{\infty}\frac{1}{(2K)^i}\sum_{\substack{j=-(i-1)\\ j\neq 0}}^{(i-1)} jP_i(a,j)\psi(a+j)$$
$$=\frac{1}{2K}\sum_{i=0}^{\infty}\frac{1}{(2K)^i}\sum_{i'=0}^{\infty}\frac{1}{(2K)^{i'}}\sum_{j=-(i'-1)}^{(i'-1)}\sum_{\substack{j'=-i\\ j+j'\neq 0}}^{+i} P_{i'}(a,j)[a+j,a+j+j']_{i+1}\psi(a+j+j'). \tag{22}$$

Proceeding as in the previous case we obtain the following relation after equating the coefficients of $\psi(a+j)$ on both sides of the equation

$$jP_i(a,j)=\sum_{i'=1}^{\infty}\sum_{j'+j=-(i-i'-1)}^{(i-i'-1)} P_{i-i'}(a,j'+j)[a+j'+j,a+j]_{i'}. \tag{23}$$

This is the recurrence relation stated in Ref. 5. There, of course, it contains a misprint in the subscript of the coefficient P which should read $(r-i)$ instead of r [cf. Eq. (3.21) of Ref. 5].

4. THE PERIODIC POTENTIAL

It has been shown[2] that the Schrödinger equation for a periodic potential, i.e., the Mathieu equation, can be written

$$\phi''+\{\lambda-2h^2\cos 2x\}\phi=0 \quad (-\pi\leqslant x\leqslant\pi),$$

where λ is real and h^2 positive.

On setting $\phi=\psi(x)e^{2h\sin x}$ and $\lambda=-2h^2+2hq+\Delta/8$, where q is (approximately) an odd integer (the latter equation incorporating the large-h asymptotic behavior of λ) this equation can be rewritten as

$$\mathscr{D}_q\psi=\frac{1}{2^5h}\left(2^4\frac{d^2}{dx^2}+2\Delta\right)\psi, \tag{24}$$

where $$\mathscr{D}_q=-2\cos x\frac{d}{dx}-(q-\sin x). \tag{25}$$

As shown in Ref. 2, the solution ψ valid for

$$|\cos(\pi/4\pm x/2)|\lesssim 1/h^{1/2}$$

can be written

$$\psi=\psi_q+\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i} P_i(q,j)\psi_{q+4j}, \tag{26}$$

where

$$\psi_q(x)=\frac{\cos^{1/2(q-1)}(\pi/4+x/2)}{\cos^{1/2(q+1)}(\pi/4-x/2)},$$

which is the solutionof $\mathscr{D}_q\psi_q=0$.

Thus ψ of (26) is a solution in terms of trigonometrical functions. It has been shown in Ref. 2 that the solutions valid in other domains have the same form as (26), with the same coefficients P_i, but with ψ_q replaced by approprite Hermite functions. Thus, the considerations given below apply also in the case of these solutions.

We now insert ψ of (26) in (24) and use the relation $\mathscr{D}_q\psi_{q+4j}=4j\,\psi_{q+4j}$. Then (see below)

$$\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i} jP_i(q,j)\psi_{q+4j}$$
$$=\frac{1}{2^7h}\{(q,q+4)\psi_{q+4}+(q,q)\psi_q+(q,q-4)\psi_{q-4}\}+\frac{1}{2^7h}\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i}P_i(q,j)$$
$$\times\{(q+4j,q+4j+4)\psi_{q+4j+4}+(q+4j,q+4j)\psi_{q+4j}+(q+4j,q+4j-4)\psi_{q+4j-4}\}, \tag{27}$$

where we have used the relation

$$\left(2^4\frac{d^2}{dx^2}+2\Delta\right)\psi_q=(q,q+4)\psi_{q+4}+(q,q)\psi_q+(q,q-4)\psi_{q-4}.$$

Here

$$(q,q+4)=(q+1)(q+3),\quad (q,q)=2[(q^2+1)+\Delta],\quad (q,q-4)=(q-1)(q-3).$$

Collecting terms in ψ_q on the right-hand side of (27), we obtain

$$\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i} jP_i(q,j)\psi_{q+4j}$$
$$=\frac{1}{2^7h}\left\{(q,q)+\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}[P_i(q,-1)(q-4,q)+P_i(q,+1)(q+4,q)]\right\}\psi_q+\left(\frac{1}{2^7h}\right)[(q,q+4)\psi_{q+4}+(q,q-4)\psi_{q-4}]$$
$$+\sum_{i=1}^{\infty}\frac{1}{(2^7h)^{i+1}}\left[\sum_{\substack{j=-i\\ j\neq 0,-1}}^{i}P_i(q,j)(q+4j,q+4j+4)\psi_{q+4j+4}+\sum_{\substack{j=-i\\ j\neq 0}}^{i}P_i(q,j)(q+4j,q+4j)\psi_{q+4j}\right.$$

$$+\sum_{\substack{j=-i\\ j\neq 0,+1}}^{i} P_i(q,j)(q+4j,q+4j-4)\psi_{q+4j-4}\Bigg]. \tag{28}$$

As before we now demand that $\{\ \}$ be zero, yielding an equation from which Δ is to be determined, i.e.,

$$0=\frac{1}{(2^7h)}\left\{(q,q)+\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\left[P_i(q,-1)(q-4,q)+P_i(q,+1)(q+4,q)\right]\right\}. \tag{29}$$

We are thus left with the following expression on the right-hand side of (28):

$$\frac{1}{(2^7h)}\{(q,q+4)\psi_{q+4}+(q,q-4)\psi_{q-4}\}+\sum_{i=2}^{\infty}\frac{1}{(2^7h)^i}\Bigg[\sum_{\substack{j=-i+1\\ j\neq 0,-1}}^{i-1}P_{i-1}(q,j)(q+4j,q+4j+4)\psi_{q+4j+4}$$
$$+\sum_{\substack{j=-i+1\\ j\neq 0}}^{i-1}P_{i-1}(q,j)(q+4j,q+4j)\psi_{q+4j}+\sum_{\substack{j=-i+1\\ j\neq 0,+1}}^{i-1}P_{i-1}(q,j)(q+4j,q+4j-4)\psi_{q+4j-4}\Bigg]. \tag{30}$$

Now, $P_i(q,0)=0$ except for $i=0$ in which case $P_0(q,0)=1$; hence the constraints "$j\neq 0$" could now be omitted. Further, since $P_0(q,j)$ is zero except for $j=0$ our expression becomes

$$\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\Bigg[\sum_{\substack{j=-i+1\\ j\neq -1}}^{i-1}P_{i-1}(q,j)(q+4j,q+4j+4)\psi_{q+4j+4}+\sum_{\substack{j=-i+1\\ j\neq 0}}^{i-1}P_{i-1}(q,j)(q+4j,q+4j)\psi_{q+4j}$$
$$+\sum_{\substack{j=-i+1\\ j\neq +1}}^{i-1}P_{i-1}(q,j)(q+4j,q+4j-4)\psi_{q+4j-4}\Bigg].$$

Thus

$$\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i}jP_i(q,j)\psi_{q+4j}$$
$$=\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\sum_{\substack{j=-i\\ j\neq 0}}^{i}\{P_{i-1}(q,j-1)(q+4j-4,q+4j)+P_{i-1}(q,j)(q+4j,q+4j)+P_{i-1}(q,j+1)(q+4j+4,q+4j)\}\psi_{q+4j}$$
$$+\sum_{i=1}^{\infty}\frac{1}{(2^7h)^i}\{-P_{i-1}(q,-i-1)(q-4i-4,q-4i)\psi_{q-4i}-P_{i-1}(q,-i)(q-4i,q-4i+4)\psi_{q-4i+4}$$
$$-P_{i-1}(q,i)(q+4i,q+4i)\psi_{q+4i}-P_{i-1}(q,-i)(q-4i,q-4i)\psi_{q-4i}$$
$$-P_{i-1}(q,i+1)(q+4i+4,q+4i)\psi_{q+4i}-P_{i-1}(q,i)(q+4i,q+4i-4)\psi_{q+4i-4}\}. \tag{31}$$

Each coefficient P_{i-1} in the second set of sums of this last expression is zero. Thus, equating coefficients of ψ_{q+4j} on both sides of the equation we obtain

$$jP_i(q,j)=P_{i-1}(q,j-1)(q+4j-4,q+4j)+P_{i-1}(q,j)(q+4j,q+4j)+P_{i-1}(q,j+1)(q+4j+4,q+4j). \tag{32}$$

This is the recurrence relation for the coefficients P given in Ref. 2. There, of course, it was written down from a consideration of all possible "moves" from i to $i-1$ in terms of permissible steps.

5. CONCLUSIONS

In the above we have demonstrated that the systematic asymptotic perturbation procedure developed in Refs. 2–5 allows a straightforward derivation of the exact recurrence relation of the perturbation coefficients of the wavefunction. Without the systematics resulting from extensive use of the recurrence relation of the unperturbed part of the wavefunction this exact relation could not be derived.

In each of the three cases discussed above we obtained an equation from which the eigenvalues can be obtained. These equations are identical with those derived previously by different methods. Since these equations have been discussed previously we do not elaborate on them here. However, the equations given here demonstrate their intimate connection with the perturbation coefficients P_i and thus with the solutions of the recurrence relations. The latter can now be treated as difference equations, and by suitable approximations their behavior for large i (i.e., large order) can be calculated. This, in turn, permits the calculation of the large-order behavior of the eigenvalue expansion as shown in Ref. 4. It is therefore of considerable interest to investigate the solutions of these recurrence relations in more detail.

ACKNOWLEDGMENTS

One of the authors (LKS) is indebted to the DAAD for the award of a visiting fellowship.

[1] R. B. Dingle, Proc. R. Soc. London Ser. A **244**, 456 (1958); **244**, 476 (1958); **244**, 484 (1958); **249**, 270 (1959); **249**, 285 (1959); **249**, 293 (1959).
[2] R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962); *ibid.* **216**, 123 (1964).
[3] H. J. W. Müller-Kirsten, Phys. Rev. D **22**, 1952 (1980).
[4] H. J. W. Müller-Kirsten, Phys. Rev. D **22**, 1962 (1980).
[5] H. J. W. Müller-Kirsten and K. Schilcher, J. Math. Phys. **9**, 255 (1968).
[6] A. F. Antippa and A. J. Phares, J. Math. Phys. **19**, 308 (1978); A. J. Phares, J. Math. Phys. **19**, 2239 (1978); P. A. Maurone and A. J. Phares, J. Math. Phys. **21**, 830 (1980).
[7] B. R. Johnson, "An expansion for power-law wave functions," J. Math. Phys. **22**, 2573 (1981).

An identity in Riemann–Cartan geometry

H. T. Nieh and M. L. Yan[a)]
Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794

(Received 24 February 1981; accepted for publication 30 October 1981)

We derive a new Gauss–Bonnet type identity in Riemann–Cartan geometry:
$\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}(R_{\mu\nu\lambda\rho}+\frac{1}{2}C^{\alpha}{}_{\mu\nu}C_{\alpha\lambda\nu})=\partial_\mu(-\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}C_{\nu\lambda\rho})$, where $C^{\alpha}{}_{\mu\nu}$ is the torsion tensor.

PACS numbers: 02.40.Ky

I. INTRODUCTION

We present an identity in Riemann-Cartan geometry which, to our knowledge does not seem to be previously known. The identity has some similarity to the Gauss-Bonnet type identities, but involves quantities which vanish when torsion is zero. With $C^{\lambda}{}_{\mu\nu}$ denoting the torsion tensor, the new identity is

$$\sqrt{-g}(\epsilon^{\mu\nu\lambda\rho}+\tfrac{1}{2}C^{\alpha}{}_{\mu\nu}C_{\alpha\lambda\rho})=\partial_\mu(-\sqrt{-g}\epsilon^{\mu\nu\lambda\rho}C_{\nu\lambda\rho}), \tag{1}$$

where the totally antisymmetric tensor density $\epsilon^{\mu\nu\lambda\rho}$ is defined according the convention

$$\sqrt{-g}\epsilon^{0123}=+1. \tag{2}$$

The identity (1) is different from the usual Gauss–Bonnet identity in that the right-hand side of (1) involves gauge-invariant quantities. It may therefore be less interesting from a topological point of view.

We first noticed the identity (1) in the context of a generalized Gauss-Bonnet identity for the de Sitter group. The identity also follows directly from the Bianchi identity with torsion. We present both derivations.

II. PRELIMINARIES

In a previous paper,[1] a simple Bianchi derivation of the Gauss-Bonnet type and Bianchi type identities is given for Riemann-Cartan geometry. We shall use the same method here to derive the identity (1), and adopt the same notations and conventions.

The basic quantites are the *vierbein* $e^{a}{}_{\mu}$ and the spin connection $\mathscr{V}^{ab}{}_{\mu}$. The matrices σ_{ab} satisfy the Lorentz algebra.

$$i\tfrac{1}{2}[\sigma_{ab},\sigma_{cd}]=\eta_{ac}\sigma_{bd}-\eta_{ad}\sigma_{bc}+\eta_{bd}\sigma_{ac}-\eta_{bc}\sigma_{ad}, \tag{3}$$

where

$$\eta_{ab}=(1,-1,-1,-1).$$

Define $R^{ab}{}_{\mu\nu}$ and $\bar{R}_{\mu\nu}$ according to

$$\bar{R}_{\mu\nu}=\tfrac{1}{4}\sigma_{ab}R^{ab}{}_{\mu\nu}\equiv\mathscr{V}_{\mu,\nu}-\mathscr{V}_{\nu,\mu}+i[\mathscr{V}_\mu,\mathscr{V}_\nu], \tag{4}$$

where

$$\mathscr{V}_\mu\equiv\tfrac{1}{4}\sigma_{ab}\mathscr{V}^{ab}{}_{\mu}. \tag{5}$$

Then

$$\begin{aligned}R^{\lambda\rho}{}_{\mu\nu}&\equiv e_a{}^{\lambda}e_b{}^{\rho}R^{ab}{}_{\mu\nu}\\&=g^{\rho\sigma}(\Gamma^{\lambda}{}_{\sigma\mu,\nu}-\Gamma^{\lambda}{}_{\sigma\nu,\mu}-\Gamma^{\lambda}{}_{\alpha\mu}\Gamma^{\alpha}{}_{\sigma\nu}\\&\quad+\Gamma^{\lambda}{}_{\alpha\nu}\Gamma^{\alpha}{}_{\sigma\mu}),\end{aligned} \tag{6}$$

where

$$\Gamma^{\lambda}{}_{\mu\nu}=e_a{}^{\lambda}(e^{a}{}_{\mu,\nu}+\mathscr{V}^{a}{}_{b\nu}e^{b}{}_{\mu}). \tag{7}$$

We can group $e^{a}{}_{\mu}$ and $\mathscr{V}^{ab}{}_{\mu}$ to form $\mathscr{V}^{AB}{}_{\mu}=-\mathscr{V}^{BA}{}_{\mu}$ (with $A=0,1,2,3,5$):

$$\mathscr{V}^{AB}{}_{\mu}:\begin{cases}\mathscr{V}^{ab}{}_{\mu}\\ \mathscr{V}^{a5}{}_{\mu}=e^{a}{}_{\mu}\end{cases}. \tag{8}$$

Making use of the de Sitter algebra as an artifice,

$$i\tfrac{1}{2}[X_{AB},X_{CD}]=\eta_{AC}X_{BD}-\eta_{AD}X_{BC}+\eta_{BD}X_{AC}-\eta_{BC}X_{AD}, \tag{9}$$

with

$$\eta_{AB}=(1,-1,-1,-1,\pm 1);$$

we define $\tilde{R}_{\mu\nu}$ and $R_{AB\mu\nu}$ according to

$$\begin{aligned}\tilde{R}_{\mu\nu}&=\tfrac{1}{4}X_{AB}\bar{R}^{AB}{}_{\mu\nu}\\&\equiv\tilde{\mathscr{V}}_{\mu,\nu}-\tilde{\mathscr{V}}_{\nu,\mu}+i[\tilde{\mathscr{V}}_\mu,\tilde{\mathscr{V}}_\nu],\end{aligned} \tag{10}$$

where

$$\tilde{\mathscr{V}}_\mu\equiv\tfrac{1}{4}X_{AB}\mathscr{V}^{AB}{}_{\mu}. \tag{11}$$

It follows from (9) and (10) that

$$\begin{aligned}\tilde{R}^{AB}{}_{\mu\nu}&=\mathscr{V}^{AB}{}_{\mu,\nu}-\mathscr{V}^{AB}{}_{\nu,\mu}\\&\quad+\eta_{CD}(\mathscr{V}^{AC}{}_{\mu}\mathscr{V}^{BD}{}_{\nu}-\mathscr{V}^{AC}{}_{\nu}\mathscr{V}^{BD}{}_{\mu}).\end{aligned} \tag{12}$$

In particular,

$$\tilde{R}^{AB}{}_{\mu\nu}=R^{AB}{}_{\mu\nu}+\eta_{55}(e^{a}{}_{\mu}e^{b}{}_{\nu}-e^{a}{}_{\nu}e^{b}{}_{\mu}), \tag{13}$$

$$\tilde{R}^{a5}{}_{\mu\nu}=e^{a}{}_{\lambda}C^{\lambda}{}_{\mu\nu}, \tag{14}$$

where $C^{\lambda}{}_{\mu\nu}$ is the torsion tensor

$$C^{\lambda}{}_{\mu\nu}=\Gamma^{\lambda}{}_{\mu\nu}-\Gamma^{\lambda}{}_{\nu\mu}. \tag{15}$$

III. PROOF OF THE IDENTITY

As has been derived in the previous paper,[1] there is the Gauss–Bonnet type identity

$$\begin{aligned}\sqrt{-g}\epsilon^{\mu\nu\lambda\rho}R^{ab}{}_{\mu\nu}R_{ab\lambda\rho}&=2\sqrt{-g}\epsilon^{\mu\nu\lambda\rho}\,\mathrm{Tr}[\tilde{R}_{\mu\nu}\tilde{R}_{\lambda\rho}]\\&=\partial_\mu\{8\sqrt{-g}\epsilon^{\mu\nu\lambda\rho}\,\mathrm{Rr}[\mathscr{V}_\nu\partial_\lambda\mathscr{V}_\rho\\&\quad-i(\tfrac{2}{3})\mathscr{V}_\nu\mathscr{V}_\lambda\mathscr{V}_\rho]\}.\end{aligned} \tag{16}$$

a)Permanent address: Department of Modern Physics, University of Science and Technology of China, Hefei, Auhui, People's Republic of China.

 0022-2488/82/030373-02$02.50 © 1982 American Institute of Physics

We can derive a similar identity for $\tilde{R}^{ab}{}_{\mu\nu}$

$$\begin{aligned}&\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\tilde{R}^{AB}{}_{\mu\nu}\tilde{R}_{AB\lambda\rho}\\&\quad=2\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\,\mathrm{Tr}[\tilde{R}_{\mu\nu}\,\tilde{R}_{\lambda\rho}]\\&\quad=\partial_\mu\{8\sqrt{-g}\,\epsilon^{\mu\nu\lambda\rho}\,\mathrm{Tr}[\tilde{\mathscr{V}}_\nu\partial_\lambda\tilde{\mathscr{V}}_\rho-i(\tfrac{2}{3})\tilde{\mathscr{V}}_\nu\tilde{\mathscr{V}}_\lambda\tilde{\mathscr{V}}_\rho]\}\end{aligned}\tag{17}$$

because

$$\epsilon^{\mu\nu\lambda\rho}\,\mathrm{Tr}[\tilde{\mathscr{V}}_\mu\tilde{\mathscr{V}}_\nu\tilde{\mathscr{V}}_\lambda\tilde{\mathscr{V}}_\rho]=0.\tag{18}$$

Subtracting (16) from (18), and taking into account (8), (13), and (14), we can obtain in a straightforward manner the identity (1).

IV. A DIRECT PROOF

For Reimann-Cartan geometry, the Bianchi identity is

$$\begin{aligned}&C^\alpha{}_{\mu\nu;\lambda}+C^\alpha{}_{\nu\lambda;\mu}C^\alpha{}_{\lambda\mu;\nu}-C^\rho{}_{\mu\nu}C^\alpha{}_{\rho\lambda}\\&\quad-C^\rho{}_{\nu\lambda}C^\alpha{}_{\rho\mu}-C^\rho{}_{\lambda\mu}C^\alpha{}_{\rho\nu}\\&\quad=R^\alpha{}_{\mu\nu\lambda}+R^\alpha{}_{\nu\lambda\mu}+R^\alpha{}_{\lambda\mu\nu},\end{aligned}\tag{19}$$

from which we obtain

$$\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}(C_{\alpha\mu\nu;\lambda}-C^\rho{}_{\mu\nu}\,C_{\alpha\rho\lambda}=\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}R_{\alpha\mu\nu\lambda}\,.\tag{20}$$

Using (15), we obtain from

$$C_{\alpha\mu\nu;\lambda}=C_{\alpha\mu\nu;\lambda}-\Gamma^\rho{}_{\alpha\lambda}C_{\rho\mu\nu}-\Gamma^\rho{}_{\mu\nu}C_{\alpha\rho\nu}-\Gamma^\rho{}_{\nu\lambda}C_{\alpha\mu\rho}\tag{21}$$

the relation

$$\begin{aligned}\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}C_{\alpha\mu\nu;\lambda}=\sqrt{-g}(&C_{\alpha\mu\nu,\lambda}-C^\rho{}_{\mu\nu}C_{\alpha\rho\nu}\\&-\tfrac{1}{2}C^\rho{}_{\alpha\lambda}C_{\rho\mu\nu})\,.\end{aligned}\tag{22}$$

Combining (20) and (22) yields

$$\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}(C_{\alpha\mu\nu;\lambda}-\tfrac{1}{2}C^\rho{}_{\alpha\lambda}C_{\rho\mu\nu})=\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}R_{\alpha\mu\nu\lambda}\,.\tag{23}$$

Since $\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}$ is constant, we then have the identity (1):

$$\begin{aligned}&\sqrt{-g}\epsilon^{\alpha\mu\nu\lambda}(R_{\alpha\mu\nu\lambda}+\tfrac{1}{2}C^\rho{}_{\alpha\lambda}C_{\rho\mu\nu}\\&\quad=\partial_\lambda(-\sqrt{-g}\epsilon^{\lambda\alpha\mu\nu}C_{\alpha\mu\nu})\,.\end{aligned}$$

We remark that the identity (1) can be directly checked without using the Bianchi identity (19).

ACKNOWLEDGMENTS

We acknowledge discussions with Professor Y. S. Wu and Professor C. N. Yang. One of the authors)M. L. Yan) wishes to thank Professor C. N. Yang for the kind hospitality at Stony Brook. This work is partially supported by the National Science Foundation under Grant No. 79-06376A01.

[1] H. T. Nieh, J. Math. Phys. **21**, 1439 (1980). There are a few misprints in this paper. Equation (16) should read as Eq. (4) of the present paper; Eq. (25) as Eq. (16) of the present paper. There should be an extra factor $\frac{1}{2}$ on the left-hand side of (30). The signs in front of the *CC* terms in the Bianchi identity (40) should be minus.

Theta functions, Gaussian series, and spatially periodic solutions of the Korteweg–de Vries equation

John P. Boyd
Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, Michigan 48109

(Received 31 July 1981; accepted for publication 4 September 1981)

It has been shown by Novikov [Funct. Anal. Appl. **8**, 236 (1974)], Dubrovin *et al.* [Russian Math. Surveys **31**, 59 (1976)], Lax [Commun. Pure Appl. Math. **28**, 141 (1975)], McKean and van Moerbeke [Inv. Math. **30**, 217 (1975)], and others that the nonlinear evolution equations which admit solitary waves also have spatially periodic exact solutions ("polycnoidal waves") which can be expressed in terms of multidimensional Riemann theta functions. Here, it is shown that via Poisson summation, the Fourier series that define the theta functions can be transformed into an infinite series of Gaussian functions. Because the lowest terms of the Gaussian series generate the usual solitary waves, it is possible to intimately explore the relationship between solitary waves and these spatially periodic "polycnoidal" waves. Also, by using the Gaussian series, one can perturbatively calculate phase velocities and wave structure for the "polycnoidal" wave even in the strongly nonlinear regime for which the soliton (or multisoliton) is the lowest order approximation. It is further shown that the Fourier series and the complementary Gaussian series both converge so rapidly in the intermediate regime of moderate nonlinearity that one may loosely state that a solitary wave is almost a linear wave, and a linear wave almost a soliton. Thus, by using both series together, one can obtain a very complete description of these stable, finite amplitude, periodic solutions. For expository simplicity, this first discussion of the Gaussian series approach to "polycnoidal" waves will concentrate on the most elementary example: the ordinary "cnoidal" wave of the Korteweg–de Vries equation. The great virtue of the Poisson method, however, is that it extends almost trivially to other equations (the Nonlinear Schrödinger equation, the Sine–Gordon equation, and a multitude of others) and also to periodic solutions of these equations that are describable in terms of higher dimensional theta functions ("polycnoidal" waves). The next to last section proves a number of generalizations of the theorems of Hirota [Prog. Theor. Phys. **52**, 1498 (1974)] applicable both to "cnoidal" and "polycnoidal" solutions without restriction, and explains how these extensions will work.

PACS numbers: 02.60. + y

1. INTRODUCTION

The first exact, nonlinear, spatially periodic solutions to an evolution equation of the class discussed here were obtained by Korteweg and de Vries[1] 85 years ago for the equation that now bears their name. They showed that their equation, henceforth referred to by the abbreviation KdV, has steadily translating waves that can be mathematically described by the elliptic cosine function $\mathrm{cn}(x; m)$. Since it was like a function whose abbreviation is "cn," they called these waves "cnoidal" waves. The nonlinear, spatially periodic solutions discussed here are generalizations of these cnoidal waves that, in the absence of any generally accepted terminology, will be referred to as "polycnoidal" waves in the rest of the paper. The reason that these generalized waves are important is that it appears that *any* spatially periodic solution of the KdV equation—or a number of other equations in the same class—can be approximated to any chosen degree of accuracy for any chosen finite time interval by an appropriate "polycnoidal" wave. Thus, to understand these generalized cnoidal waves is also to understand the general spatially periodic solution to the Korteweg–de Vries and other evolution equations.

The elliptic cosine depends on a parameter m (the "modulus"); Korteweg and de Vries showed that the limit $m = 0$ corresponds to a linear wave and the elliptic cosine reduces to the ordinary cosine. In the limit $m \to 1$, the spatial period of the wave becomes infinite, the elliptic cosine becomes the hyperbolic secant function, and the cnoidal wave becomes the solitary wave discovered observationally in 1831[2]: an isolated, steadily translating, finite amplitude peak of permanent form. Equivalently, if one rescales the variables so as to keep the spatial period fixed, the solitary wave or "soliton" corresponds to the limit of infinitely large amplitude (the spatial structure of the soliton tends toward that of a delta function) while the linear wave as usual is the limit of infinitesimal amplitude.

"Polycnoidal" waves also tend to linear waves and solitons in appropriate limits, and the relationship between these limits and the actual polycnoidal waves of intermediate amplitude is one of the major themes of this paper. Before discussing how we propose to explore these relationships, it is appropriate (and necessary) to briefly review the major developments in the theory of the Korteweg–de Vries equation.

Most work on this equation has studied its solutions subject to two different species of initial/boundary conditions: (i) the unbounded problem in which $x \in [-\infty, \infty]$ and the initial condition is localized, i.e., is exponentially small everywhere outside of a finite interval, and (ii) the spatially periodic problem in which both the initial condition and the

 0022-2488/82/030375-13$02.50 © 1982 American Institute of Physics

solution for all later times are required to be periodic functions of x. Korteweg and de Vries were able to derive a solution for the unbounded problem (the solitary wave) as a limit of their class of solutions for the periodic problem, but in later work, these two problems have represented completely separate lines of development.

Little progress with either was made until 1967, when Gardner *et al.*[3] showed the unbounded problem, although the KdV equation itself is nonlinear, could be solved exactly through a sequence of solving only linear equations: the so-called inverse scattering method. This procedure was to the study of nonlinear equations what the Rosetta Stone was to Egyptology, and it was subsequently extended to a large class of other equations, including the Nonlinear Schrödinger equation, the sine–Gordon equation, and many others. Although for simplicity, the discussion here will concentrate on the KdV equation, the ideas and techniques explained here extend to all members of this class of "exactly integrable" nonlinear equations.

The inverse scattering analysis showed that the general solution to the KdV equation (and its fellows) consists of two parts: a finite number of solitary waves or "solitons," which are permanent, isolated, finite amplitude waves, plus "radiation," which is used as a catch-all term to describe the miscellaneous peaks and ripples that eventually disperse so that the solitons are the sole asymptotic solution as $t \to \infty$. The reason for the name "soliton," with its connotation of particlelike rather than wavelike properties, is that when solitons collide, they eventually emerge from collision unchanged in shape, size, or speed except for a phase-shift.

Unfortunately, although inverse scattering is so useful for theoretical and qualitative purposes, researchers have found that for obtaining quantitative results and case studies, the steps of the inverse scattering method are so cumbersome that it is easier to numerically integrate the KdV equation directly using a conventional time-marching scheme. Because of this clumsiness, Hirota[4,5] introduced an alternative approach to the unbounded problem which is the direct ancestor of the methods to be used here.

Hirota's technique is based on a logarithmic transformation of the dependent variable to give a transformed version of the KdV equation which, to avoid confusion, will be referred to as the Hirota–Korteweg–de Vries equation or H–KdV for short. Hirota showed that although the H–KdV equation is nonlinear, it is possible to construct exact solutions by adding finite sums of exponentials from which the exact, multisoliton solutions of the KdV equation can be obtained through the inverse transformation. Hirota's method is very simple and involves nothing more exotic than differentiation, logarithms, and exponentials, but it has the weakness of excluding the "radiation" part of the general KdV solution. However, for smooth, large-amplitude initial conditions, almost all of the initial energy goes into the solitons anyway (referred to as the principle of "soliton dominance" in Boyd[6]), so this restriction is not fatal, and Hirota's method is still actively used in research today (for example, Ma and Redekopp[7]) even when the exact inverse scattering algorithm is known for the same equation (Ma[8]).

Meanwhile, independent groups of American[9] and Russian[10,11] mathematicians have developed an analog of the inverse scattering method for obtaining a class of generalizations of the cnoidal waves for (ii), the spatially periodic problem. Unfortunately, it is again true that the formal solution, which will be here called the "Hill's spectrum method," is as cumbersome computationally as it is powerful theoretically.

However, the author and Nakamura[12] independently realized that there is an alternative. The transformation that relates the theta functions to the actual solution of the Korteweg–de Vries equation is exactly that made by Hirota—in other words, the theta functions are exact solutions not of the KdV equation itself but rather of Hirota's transformed version, the H–KdV equation. This suggests what Nakamura calls the "direct method": computing nonlinear phase speed corrections and theta function parameters via direct substitution of the theta series into the H–KdV equation and matching of Fourier series coefficients. Thus, there are now two approaches—the formal exact method based on quantum mechanics potential theory and a more heuristic but vastly simpler alternative based on solving Hirota's transformed equations—for both the infinite and periodic spatial domains.

In later work, Nakamura, Hirota, Ito, and Matsuno[13–15] have greatly extended this direct theta function procedure and their papers are highly original and a treasurehouse of useful information. However, they work exclusively with the Fourier series representation of the theta functions and omit all mention of what are the principal themes of this paper: that a direct method using the Gaussian series representation is not only possible but is more useful than its Fourier series counterpart.

The reason for the greater usefulness of the Gaussian series is that one can obtain most of the information derivable from the theta–Fourier series by directly attacking the Korteweg–de Vries equation via the singular perturbation technique known variously as the method of strained parameters or the method of multiple scales.[16,17] Because the multiple-scales approach is conceptually useful also in understanding precisely what a "polycnoidal" wave is, some multiple scales calculations for polycnoidal waves are given in Appendix B.

The Gaussian series representation, however, is directly tied to the fact that the exact solutions of the H–KdV equation are theta functions, and it gives results that cannot be reproduced by any conventional perturbation scheme. As a consequence, the rest of the paper will concentrate upon the Gaussian representation of the theta function, except for the appendices.

Sections 2–5 describe the simplest example: the ordinary cnoidal wave for the Korteweg–de Vries equation. Despite the fact that this was solved 85 years ago, the theta–Gaussian method will nonetheless yield some new results. Later sections of the paper and future work now in progress will deal with polycnoidal waves and other evolution equations. But the point in beginning with this simple example is that the ideas explained through it are the key to everything else.

2. THE HIROTA–KORTEWEG–DE VRIES EQUATION AND THE THETA FUNCTION

The Korteweg–de Vries equation itself is

$$u_t + uu_x + u_{xxx} = 0. \tag{2.1}$$

The transformation to the H–KdV equation is made in two stages. First, set $u = p_x$ and then integrate (2.1) once with respect to x to obtain

$$p_t + 1/2p_x^2 + p_{xxx} = A, \tag{2.2}$$

where A is an arbitrary constant of integration. In deriving the multiple soliton solutions, one can set $A = 0$ without loss of generality, but as noted by Nakamura,[12] this is absolutely disastrous for the polycnoidal wave. Instead, A must be computed as a function of the parameters in the same way as the nonlinear phase speed. The second step is to introduce the nonlinear transformation

$$p = 12(\ln F)_x. \tag{2.3}$$

Substituting (2.3) in (2.2), on finds that all third and fourth degree terms in F identically cancel to leave the H–KdV equation:

$$F(F_t + F_{xxx})_x - F_x(F_t + F_{xxx}) + 3(F_{xx}^2 - F_xF_{xxx}) = AF^2. \tag{2.4}$$

Hirota showed that (2.4) and indeed all his transformed equations can be expressed more compactly by using certain bilinear operators, but this alternative version of (2.4) will be deferred until Sec. 6, where Hirota's bilinear operators will be useful in proving certain theorems. The solution of the KdV equation is

$$u = 12(\ln F)_{xx}, \tag{2.5}$$

where F is a theta function.

The argument of the theta function is

$$X = x - ct + \phi, \tag{2.6}$$

where c is the phase speed and ϕ a constant phase factor. For higher-dimensional theta functions, we have additional arguments of the form $Y = k_2(x - c_2t) + \phi_2$, $Z = k_3(x - c_3t) + \phi_3$, as one would anticipate from the fact that the polycnoidal waves reduce to multiple solitary waves in the appropriate limit. Since the N-soliton solution consists of N distinct peaks each with its own width and phase speed, it follows that a function of N variables of the form of X, Y, Z, etc., are necessary to provide a compact description of the waves. By rescaling via a theorem proved in Sec. 7, one can always set one of the wave numbers k equal to 1 without loss of generality, and this has been done in (2.6).

For the one-dimensional case, one can take the theta function to be

$$\theta_4(X; q) = 1 + \sum_{n=1}^{\infty} (-1)^n q^{n^2} \cos(2nX), \tag{2.7}$$

where q is a constant known as the "nome." When working with theta functions, the nome is a more convenient measure of the "ellipticity" of the elliptic cosine than the modulus m; the relationship between them is given in Appendix A. However, q has one thing in common with m: the limit $q \to 0$ again gives the linear wave while $q \to 1$ gives the solitary wave. From the form of the Fourier series in (2.7), one can show that it converges uniformly and absolutely for all $q < 1$, but it is self-evident that the rate of convergence becomes poorer and poorer as the soliton limit is approached. Consequently, it is the alternative Gaussian series for the theta function that one must use to explore the relationship between solitons and cnoidal waves.

Nonetheless, even from a perspective focused entirely on Fourier series, the theta series (2.7) converges much more rapidly than that for $u(x)$ itself, which is

$$u = 12(\ln \theta_4)_{XX} = 96q \sum_{n=1}^{\infty} \frac{nq^{n-1}}{1 - q^{2n}} \cos(2nX), \tag{2.8}$$

as obtained by differentiating the known[18] series for the logarithmic derivative of the theta function. The elliptic cosine, whose square gives $u(x)$ (see Appendix A for details), is a meromorphic function, and this alone[19] is enough to prove that the coefficients of the Fourier series must be asymptotically $O(q^n)$ for some constant q, $|q| < 1$. [This, in fact, is true for the hyperelliptic functions that give $u(x, t)$ for the polycnoidal waves as well.] The theta functions, however, whether in one or many dimensions, are entire functions. For θ_4, we see that the Fourier coefficients are $O(q^{n^2})$ so that 10 terms of the theta Fourier series give the same accuracy as 100 terms of the Fourier series for cn^2. It is precisely this very rapid convergence for the entire function as opposed to the meromorphic function that led C. G. Jacobi to introduce the theta functions in the first place and build his entire approach to elliptic functions around them. It is precisely because the Gaussian series for the theta functions shares this same very rapid convergence that it will be shown to be a powerful tool for understanding strongly nonlinear, spatially periodic waves.

Poisson summation[20] of (2.7) gives the Gaussian series, which may be written in either of the two forms:

$$\theta_4(X; q) = \begin{cases} 2s^{1/2}\exp\left\{-sX^2/\pi\right\} \sum_{n=0}^{\infty} q'^{(n+1/2)^2}\cosh[(2n+1)sX] & (2.9) \\ s^{1/2} \sum_{-\infty}^{\infty} \exp[-s(X - \tfrac{1}{2}\pi(2n+1))^2/\pi]. & (2.10) \end{cases}$$

where q', the "complementary nome," and s are defined by

$$q' \equiv e^{\pi^2/\ln q}, \tag{2.11}$$

$$s = -\pi/\ln q. \tag{2.12}$$

The relationships between q, q', and s can be expressed more symmetrically in the form

$$q = e^{-\pi/s}, \tag{2.13}$$

$$q' = e^{-\pi s}. \tag{2.14}$$

As $q \to 1$, $q' \to 0$ and vice versa so that the Gaussian series [(2.9) or (2.10)] and the Fourier series (2.7) are indeed complementary, with one converging rapidly in the parameter range where the other converges slowly.

The first form of the Gaussian series, (2.9), which can be obtained from (2.10) by multiplying out the exponents in (2.10), extracting the common factor of $\exp(-sX^2/\pi)$, and combining exponentials into hyperbolic cosines, is the one that is most closely analogous to the Fourier series. Because

q' appears explicitly in (2.9), this form is the one that is useful for practical perturbative calculations.

The second form, (2.10), is more useful conceptually: it shows that the theta function can be represented as a series of Gaussians of identical size and shape spaced at intervals of π on the whole interval $X\in[-\infty, \infty]$. For large q' (small q), the Gaussians overlap heavily and the shape is roughly that of an ordinary cosine function with symmetrical crests and troughs. For small q'—the near-soliton-limit—the Gaussians are well separated and the theta function takes the form of sharp, narrow peaks separated by broad, flat troughs.

In the limit $q'\to 0$, only the two Gaussians at $X = \pm \pi/2$ are significant because the contributions of all the other terms in the series are exponentially small on the interval $X\in[-\pi/2, \pi/2]$. This suggests that—despite all appearances to the contrary—this limiting solution of two Gaussians is somehow equivalent to the single soliton solution of Hirota, which is the sum of a constant plus an exponential. In the next section, we shall see that this is indeed the case and explore how the transition from soliton to cnoidal wave is made.

3. THE BI-GAUSSIAN SOLITON AND THE PROBLEM OF PERIODICITY

A. The bi-Gaussian soliton

By direct substitution, one can show

$$\Theta(X) = e^{-(L/2)X^2}\cosh(sX) \tag{3.1}$$

is an *exact* solution of the Hirota–Korteweg–de Vries equation (2.4) for *arbitrary* values of s and L provided that

$$c = 4s^2 - 12L, \tag{3.2}$$

$$A = 2L(2s^2 - 3L), \tag{3.3}$$

where c is the phase speed, A the constant of integration in the H–KdV equation, and $X = x - ct + \phi$ as defined by (2.6) with ϕ arbitrary. $\Theta(X)$ is[21] the sum of two Gaussians with peaks at $X = \pm s/L$, so it will be referred to as the bi-Gaussian soliton solution. Taking the second logarithmic derivative gives

$$u(x,t) = -12L + 12s^2\,\mathrm{sech}^2[s(x - ct)]. \tag{3.4}$$

As mentioned earlier, the theta functions of the proceeding section are functions of but a single parameter—and so are the one-soliton solutions of the KdV equation in the form usually given—whereas the bi-Gaussian $\Theta(X; s, L)$ contains *two* independent parameters. However, one can see from (3.4) that the extra parameter corresponds merely to the freedom (mathematically, if not physically!) to change mean sea level in the Korteweg–de Vries equation. In mathematical terms, it is trivial to prove that if $u(x - ct)$ is any solution to the KdV equation, than

$$\tilde{u} \equiv a + u(x - (c + a)t) \tag{3.5}$$

is also a solution. Thus, cnoidal waves and the one-soliton solution are really two-parameter families, but one of the parameters, the constant a in (3.5), is trivial, and only the nome q is significant. However, with the bi-Gaussian, we can obtain the full two-parameter single-soliton family by varying s and L.

Hirota's solution is the special case $L = 0$ that gives

$$\Theta = \cosh(sX). \tag{3.6}$$

This gives the usual soliton with zero mean sea level, but it is not quite in standard form itself. However, one can show[22] that any solution of the H–KdV equation is still a solution if multiplied by an arbitrary constant times an exponential whose argument is linear in X. Multiplying (3.6) by $2\exp(sX)$ gives

$$F = 1 + \exp(2sX), \tag{3.7}$$

which is the form in which Hirota's 1-soliton solution is usually given.

Unfortunately, to extend this bi-Gaussian function into the infinite series for the cnoidal wave, we must take a nonzero value of L. This in turn means accepting the annoying complication of a shift in mean sea level via the term $-12L$ in (3.4) and an identical shift also in the phase speed (3.2). To have generalized Hirota's one-parameter solution (3.7) to a two-parameter one is not in itself a very useful accomplishment. However, it is the bi-Gaussian—not (3.7)—that generates the cnoidal wave, so one must understand the shifts introduced by L to make correct comparisons between exact and approximate cnoidal wave solutions as shall be done in Sec. 5.

B. Periodicity

If one approached the problem of computing approximate periodic solutions to the KdV equation with no knowledge of theta functions—but a knowledge of solitons—one heuristic approach would be to approximate the cnoidal wave by an infinite series of evenly spaced hyperbolic secant functions, i.e.,

$$u(x, t) \doteq 12s^2 \sum_{n=-\infty}^{\infty} \mathrm{sech}^2(s(X - n\pi)). \tag{3.8}$$

When s is large, the soliton peaks are narrow and well-separated with little overlap, so (3.8), with c given by the usual soliton formula (3.2), is indeed a consistent first approximation to the cnoidal wave with an error which decreases exponentially fast as s increases.

The only problem is that there is no particularly good way to calculate higher-order corrections—to explore precisely how the periodicity has altered the cnoidal wave from the soliton. One could substitute (3.8) into the Korteweg–de Vries equation, but the inhomogeneous terms even at lowest order would involve fourth powers of reciprocal hyperbolic functions, and inverting the linear part of the Korteweg–de Vries equation requires inverting a partial differential operator. Furthermore, the error in neglecting higher values of n in the series (3.8) is an exponential function of s, but algebraic powers of s would also appear. In short, perturbative theory using (3.8) as the lowest approximation would be a horrible mess, requiring great analytical ingenuity to obtain even the first and second corrections. Furthermore, the shape of u as a function of X, as well as the nonlinear phase speed c, would both have to be corrected order-by-order.

If we apply this same heuristic philosophy to the

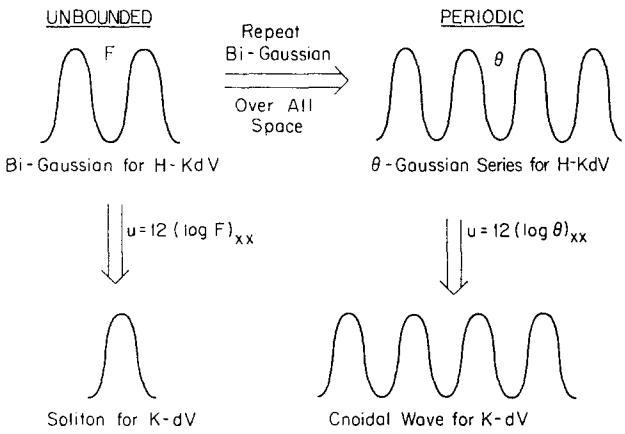

FIG. 1. Schematic diagram showing the relationship between the bi-Gaussian and theta functions solutions to the H–KdV equation. The left side shows the situation when the domain is unbounded: the solution to the H–KdV equation has just two peaks on all of $X\in[-\infty, \infty]$, and the second logarithmic derivative of this gives a single peak which is the usual solitary wave. When bi-Gaussian pattern is repeated with even spacing over all X, it generates the Gaussian series of the theta function. This, as shown on the right, is a periodic solution of the H–KdV equation and its second logarithmic derivative gives the usual cnoidal wave.

H–KdV equation, using the bi-Gaussian soliton (3.1) with

$$L = (2/\pi)s, \tag{3.9}$$

so that the peaks of the Gaussians are π units apart, we obviously obtain an infinite series of evenly spaced Gaussians as indicated schematically in Fig. 1. But we have already seen in the previous section that such a series of Gaussians is an *exact* representation of the theta function. Consequently, by taking the single soliton solution and repeating it with even spacing over the whole spatial domain, we obtain the *exact* solution for the spatially periodic problem—but only when we work through the H–KdV equation, Hirota's transformed equation, rather than through the KdV equation itself. The series of hyperbolic secant functions in (3.8) is only approximate.

The same strategy works equally well to generate periodic extensions of the multiple soliton solutions. The double soliton, for example, is given by the sum of four evenly spaced Gaussians forming a square in the XY plane where $X = x - c_1 t + \phi_1$, $Y = k_2(x - c_2 t) + \phi_2$, and where each Gaussian is now an exponential whose argument is a second-degree polynomial in both X and Y. Repeating this basic four-Gaussian unit over the whole XY plane with even spacing gives the Gaussian series for the two-dimensional theta function, and the second logarithmic derivative of this gives the double cnoidal wave solution to the KdV equation.

Because only the nonlinear phase speed (or speeds) need be corrected, it is trivial in principle (although the algebra can become tedious) to calculate the cnoidal and polycnoidal waves to any order. For the cnoidal wave, for example, one obtains at each order two linear equations in two unknowns which can be solved to obtain c and A to that order, with the expansion proceeding in powers of the complementary nome q': in Sec. 4, we shall show how one can obtain the full infinite series for c almost trivially.

Before we do this, however, one final point must be made; Hirota's single soliton solution for the H–KdV equation, (3.7), is spatially *unbounded* even though, thanks to the magic of the logarithmic derivative, it gives the usual sech^2 solution when the transformation back to the KdV equation is made. Consequently, it is not possible to add evenly spaced Hirota functions of the form of (3.7) to obtain any sort of meaningful series. It was therefore absolutely necessary to generalize Hirota's solution to the bi-Gaussian so as to obtain a form for the single soliton which would be *spatially bounded* and *localized* for the Hirota–Korteweg–de Vries equation, too. Only then would it be possible to repeat the single-soliton solution over the whole domain with even spacing to obtain a solution that is manifestly periodic and solves the H–KdV equation exactly.

It is the bi-Gaussian, and not the simple exponential solution of Hirota, that is the proper generator of the cnoidal wave.

4. THE RESIDUAL EQUATIONS AND THE EXACT SOLUTION FOR THE CNOIDAL WAVE

A. The residual equations

The first step in obtaining the full series solution is to rewrite the Gaussian series (2.9) as

$$\theta_4(X; q) = s^{1/2} e^{-sX^2/\pi} \sum_{\substack{-\infty \\ \text{[half integers]}}}^{\infty} q'^{n^2} e^{2nsX}, \tag{4.1}$$

where the hyperbolic cosines have been broken up into pairs of exponentials and where, in order to eliminate factors of $\frac{1}{2}$, the sum is taken over all the "half integers" $\{\ldots, -\frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots\}$. Substituting this into the H–KdV equation gives the residual

$$\rho \equiv s e^{-2sX^2/\pi} \sum_{\substack{-\infty \\ \text{[half integers]}}}^{\infty} \sum_{-\infty}^{\infty} q'^{n^2+n'^2} \zeta(n - n', c, A) e^{2(n+n')sX}. \tag{4.2}$$

We must solve for c and A such that $\rho \equiv 0$.

One might expect that, since the H–KdV equation (2.4) involves differentiations of up to fourth order with respect to x and first order with respect to time, the Gaussian in (4.1) would cause $\zeta(n - n'; c, A)$ to be a polynomial of fourth degree in x and first degree in t. In fact, as shall be proven in Sec. 6, because of cancellations ζ is *independent* of both x and t, just as if we had substituted the Fourier series for θ_4 into the H–KdV equation instead.

The second important property of ζ—also noted by Nakamura for its Fourier series equivalent—is that it is a function only of the *difference* $(n - n')$, and not of n and n' separately. This property, again proved in Sec. 6, is true for all of Hirota's transformed nonlinear evolution equations, and is sufficient—without specification of the precise form of $\zeta(n - n'; c, A)$—to prove that the residual is the sum of two theta functions. Consequently, the exact definition of ζ will be deliberately postponed to the next subsection.

Keeping these two properties of ζ in mind, the next step is to define $j = n + n'$ and collect terms in $\exp(2jsX)$ to obtain

$$\rho = s e^{-2sX^2/\pi} \sum_{\substack{j=-\infty \\ \text{[integers]}}}^{\infty} R_j e^{2jsX}, \tag{4.3}$$

where

$$R_j \equiv \sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{n^2+(j-n)^2}\zeta(2n-j;c,A). \tag{4.4}$$

The residual ρ vanishes if and only if all the $R_j = 0$, where j is any integer, so we seem to have rather more equations (an infinite number!) than we have unknowns (2, c and A). However, it is easy to prove that all even R_j are multiples of R_0 and all the odd R_j are multiples of R_1.

For brevity, we shall deal only with the even R_j. From (4.4)

$$R_{2J} \equiv \sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{n^2+(n-2J)^2}\zeta(2[n-J]). \tag{4.5}$$

Defining

$$N \equiv n - J \tag{4.6}$$

gives

$$R_{2J} = \sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{(N+J)^2+(N-J)^2}\zeta(2N) \tag{4.7}$$

$$= q'^{2J^2} \sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{2N^2}\zeta(2N) \tag{4.8}$$

$$= q'^{2J^2}R_0 \quad \text{for all integers } J. \tag{4.9}$$

Similarly, one can show

$$R_{2J+1} = q'^{2J^2+2J}R_1. \tag{4.10}$$

By using (4.1) and the equivalent series for $\theta_3(x;q)$ $[\equiv\theta_4(x+\pi/2;q)]$, which is identical in form with (4.1) except that the sum is taken over all integers instead of half integers, one can rewrite (4.2) as

$$\rho = [R_0\theta_3(2X;q'^2) + R_1\theta_4(2X;q'^2)]se^{-sX^2/\pi}. \tag{4.11}$$

Thus, the solution of the cnoidal wave has been reduced to solving two equations in two unknowns

$$R_0(c,A) = 0, \tag{4.12}$$

$$R_1(c,A) = 0. \tag{4.13}$$

For the Korteweg–de Vries equations, these coupled equations are *linear*; for the Boussinesq equation of Nakamura,[12] these equations are quadratic in c, but can still be solved explicitly.

B. Theta matrix doubling

One important aspect of (4.11) that is independent of the nonlinear evolution equation is that all the quantities involved—R_0, R_1, the theta functions, c, and A—are functions of q'^2 rather than q'. In particular, the series that define R_0 and R_1, are series in q'^{2n^2}, and thus converge much more rapidly than even the fast-converging series for the theta functions themselves, whose coefficients are q'^{n^2}. This same phenomenon holds for the polycnoidal waves, too, where it will be called "theta matrix doubling" for notational reasons explained in Sec. 7. It is this very rapid convergence that makes perturbative calculations for two- and three-dimensional theta functions feasible.

C. The exact solution for the Korteweg–de Vries equation

For the Korteweg–de Vries equation, one can show by direct substitution of (4.1) into the H–KdV equation (3.4) or by using general theorems of Hirota's bilinear operations proved in Sec. 6 that

$$\zeta(n-n';c,A) = \left\{\frac{48s^2}{\pi^2} - 2A + \frac{4sc}{\pi}\right\} + 16s^4(n-n')^4 - 4cs^2(n-n')^2 - \frac{96s^3}{\pi}(n-n')^2, \tag{4.14}$$

which depends on n and n' via

$$\Delta = (n-n')^2, \tag{4.15}$$

and (4.12) and (4.13) become the matrix equation

$$\begin{vmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{vmatrix} \begin{vmatrix} c \\ A \end{vmatrix} = \begin{vmatrix} \Omega_1 \\ \Omega_2 \end{vmatrix}, \tag{4.16}$$

where

$$\Lambda_{11} = \frac{4s}{\pi}\left(\sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{2n^2}\right) - 16s^2\left(\sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} n^2 q'^{2n^2}\right), \tag{4.17}$$

$$\Lambda_{12} = -2\left(\sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{2n^2}\right), \tag{4.18}$$

$$\Omega_1 = 16s^2 \sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} \left(\frac{3}{\pi^2} - \frac{24s}{\pi}[n^2] + 16s^2[n^4]\right)q'^{2n^2}, \tag{4.19}$$

and where the elements of the second row (corresponding to R_1) are, after multiplication by $q'^{1/2}$, identical with those of the first row except that (i) the sums are taken over the *integers* instead of the half integers and (ii) the terms $n = 0$ are taken with a factor of $\frac{1}{2}$ as in (4.15). A more compact description can be obtained by defining

$$H(q') \equiv \sum_{\substack{n=-\infty \\ [\text{half integers}]}}^{\infty} q'^{2n^2} = \frac{-\pi^{1/2}}{[\ln q]^{1/2}}\theta_4(0;q^{1/2}), \tag{4.20}$$

$$I(q') \equiv \sum_{\substack{n=-\infty \\ [\text{integers}]}}^{\infty} q'^{2n^2} = \frac{-\pi^{1/2}}{[\ln q]^{1/2}}\theta_3(0;q^{1/2}), \tag{4.21}$$

where the right-hand sides of (4.20) and (4.21) follow by evaluating (4.1) and the corresponding series for θ_3 at $X = 0$ and then using the usual relationship between q and q', (2.11). The fact that q appears as $q^{1/2}$ is a consequence of the fact that the series defining $H(q')$ and $I(q')$ are theta-function series in q'^2 rather than q' itself as explained in the previous subsection. Letting a subscripted q' denote differentiation with respect to q', we have

$$\Lambda_{11} = (4s/\pi)H - 8s^2q'H_{q'}, \tag{4.22}$$

$$\Lambda_{12} = -2H, \tag{4.23}$$

$$\Omega_1 = 16s^2\left(\frac{3H}{\pi^2} - \frac{12sq'}{\pi}H_{q'} + 4s^2q'[q'H_{q'}]_{q'}\right), \quad (4.24)$$

$$\Lambda_{21} = \frac{4s}{\pi}I - 8s^2q'I_{q'} - \frac{2s}{\pi}, \quad (4.25)$$

$$\Lambda_{22} = -2I + 1, \quad (4.26)$$

$$\Omega_2 = 16s^2\left(\frac{3}{\pi^2}I - \frac{12sq'}{\pi}I_{q'} + 4s^2q'[q'I_{q'}]_{q'}\right) - \frac{24s^2}{\pi^2}, \quad (4.27)$$

where s and q' are related via $q' = \exp(-\pi s)$.

One finally obtains

$$c = \frac{(\Omega_1\Lambda_{22} - \Omega_2\Lambda_{12})}{\Lambda_{11}\Lambda_{22} - \Lambda_{12}\Lambda_{21}} \quad (4.28)$$

and a similar expression for A, which is not of physical interest.

5. COMPARISON OF THE EXACT AND APPROXIMATE SOLUTIONS

The two series representations of the theta function are complementary with the Fourier series converging more rapidly for small q and the Gaussian series more rapidly for small q'. The worst possible case is when

$$q = q', \quad (5.1)$$

because then both converge equally well or badly. Consequently, we can limit our attention to this single, "worst case" value of q' which is

$$q' = e^{-\pi} = 0.0432. \quad (5.2)$$

The astonishingly small value of q' makes one feel like cheering. As noted earlier, the theta series converge much faster than an ordinary geometric series because the terms are proportional to q'^{n^2} rather than q'^n, but q' is so small that even the lowest-order approximation is very accurate. The approximate solution to $O(q'^2)$ is

$$u = -24s/\pi + 12\,\mathrm{sech}^2(sX)[1 - 8q'^2\cosh^2(sX) + 16q'^2\cosh^4(sX)] + O(q'^4), \quad (5.3)$$

$$c = -24s/\pi + 4s^2 - 96s^2q'^2, \quad (5.4)$$

where $q' = \exp(-\pi s)$ with

$$s = 1 \text{ at } q = q' = 0.0432. \quad (5.5)$$

The corresponding exact solution in terms of elliptic function is (at $s = 1$)

$$u = -3.81973 + 8.35918cn^2(1.1804X; m = \tfrac{1}{2}), \quad (5.6)$$

$$c = -3.81973, \quad (5.7)$$

where m is the modulus.[23] Although we have been mostly concerned with the Gaussian series, the Fourier series approximations are still sufficiently interesting to be included in our comparisons and are

$$u \doteq 96q\cos(2X) + 192q^2\cos(4X) + O(q^3), \quad (5.8)$$

$$c \doteq -4 + 96q^2, \quad (5.9)$$

where the first follows from (2.8) and the second is derived in Appendix A.

The lowest-order and second-lowest-order approximations to $u(x, t)$ are compared in Figs. 2 and 3. The agreement between the exact and approximate graphs is remarkable;

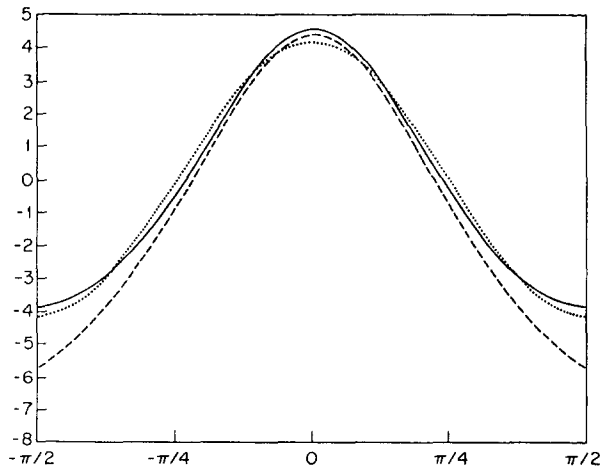

FIG. 2. A comparison of the lowest Fourier approximation (dotted line, a simple cosine function) and the lowest Gaussian series-derived approximation (dashed line, a constant plus the hyperbolic secant squared) with the exact cnoidal wave for $q = q' = 0.0432$.

the lowest-order approximations would be acceptable for most purposes and the second-order approximations are almost indistinguishable from the exact solution.

Similar remarks apply to phase speeds. The lowest-order approximations and their errors are

$$c^{(1)}_{\text{Gaussian}} = -24/\pi + 4 = -3.639, \quad (5.10)$$

Absolute error = 0.181,

Relative error = 4.7%;

$$c^{(1)}_{\text{Fourier}} = -4, \quad (5.11)$$

Absolute error = −0.180,

Relative error = −4.7%;

the second-order approximations are

$$c^{(2)}_{\text{Gaussian}} = -3.81871, \quad (5.12)$$

Absolute error = 0.00102,

Relative error = 0.027%;

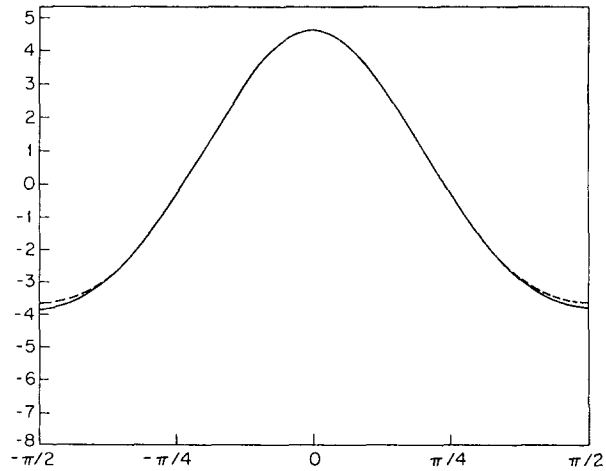

FIG. 3. A comparison of the second Gaussian approximation (5.3) with the exact cnoidal wave for the "worst case" $q = q' = 0.0432$. The second Fourier approximation is indistinguishable from the exact solution to within the thickness of the curve.

$$c^{(2)}_{\text{Fourier}} = -3.8207, \qquad (5.13)$$
$$\text{Absolute error} = -0.00100,$$
$$\text{Relative error} = -0.026\%.$$

It is amusing that even in this weakly nonlinear regime—note that $c^{(1)}_{\text{Fourier}}$ is accurate to better than 5%—the solitary wave formula (representing the strongly nonlinear regime) is just as accurate. Thus, the two regimes strongly overlap. It is in this sense that one can say that the solitary wave is "almost a linear wave": it yields an excellent approximation even when q is small enough so that the lowest-order Fourier approximation—a simple cosine function—also gives an accurate approximation. In the same sense in reverse, one can say that the linear wave is "almost a soliton" in that it gives an accurate approximation when the nonlinearity is so strong the wave shape and speed are also accurately given by those of the solitary wave.

This strong overlapping of the linear and solitary wave regimes has two important implications. First, it suggests that perturbation theory will yield useful, understandable results for the "polycnoidal" wave also. Obviously, if one needed to carry the expansion to high order in N different parameters, perturbation theory would be pointless, and one would learn as much—or as little—with much less work by staring at films of numerical integrations of the KdV equations. The accuracy and overlap of the expansions for the ordinary cnoidal wave suggest that this will not be the case; suggest instead that the lowest- or second-lowest-order perturbation theory will be more than adequate.

The other implication is conceptual. The phrase "solitary wave" has the obvious connotation of a single, isolated wave peak. What has been shown here, however, is that a wave that to the eye looks like an ordinary linear cosine function—and as Fig. 2 shows, is well approximated by a cosine function—may nevertheless be accurately modeled by a solitary wave. The isolation of a wave or a wave peak from its fellows is not an essential ingredient either in the balance between nonlinearity and dispersion, which allows the solitary wave and cnoidal wave to exist as stable, permanent forms, or in the mathematical approximation of the wave by the characteristic sech^2 shape and speed of the soliton. Thus, the intuitive equivalence of "solitary" with "isolated" has been shown here to obscure the fact that such an isolated peak and a not-very-steep cnoidal wave are essentially the same thing.

6. HIROTA'S BILINEAR OPERATORS AND SOME THEOREMS ABOUT THEM

Hirota[4] showed that his transformed nonlinear evolution equations could always be expressed in terms of the bilinear operators defined by

$$D_x^n D_t^m (F\cdot G) \equiv \left[\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^n \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^m F(x,t)G(x',t')\right]_{\substack{x'=x \\ t'=t}} \qquad (6.1)$$

where the notation indicates that x' and t' are to be replaced by x and t after the differentiations have been performed. The Hirota–Korteweg–de Vries equation, for example, is

$$(D_x^4 + D_x D_t)(F\cdot F) = 2AF^2 \qquad \text{[H-KdV]}, \qquad (6.2)$$

which is completely equivalent to (2.4).

Hirota proved a great many theorems and corollaries about the action of these bilinear operators on exponentials with linear arguments. In particular

$$D_x^n D_t^m (e^{(kx+wt)}\cdot e^{(k'x+w't)}) = (k-k')^n (w-w')^m e^{(k+k')x+(w+w')t}. \qquad (6.3)$$

Notice that the result depends only on the *difference* between k and k' and between w and w'. When a theta function Fourier series such as

$$\theta_3 = \sum_{\substack{n=-\infty \\ [\text{integers}]}}^{\infty} q^{n^2} \exp[2\pi i n(kx+wt)] \qquad (6.4)$$

is substituted into (6.2), the result is to generate a doubly infinite series of cross terms of the form of (6.3) so that the residual is

$$\bar{\rho} = \sum_{n=-\infty}^{\infty} \sum_{n'=-\infty}^{\infty} q^{n^2+n'^2} \bar{\zeta}(n-n'; c, A) e^{2\pi i(n+n')X}, \qquad (6.5)$$

with $X = kx + wt$, which is identical[24] in form to (4.2) except that it is a Fourier series instead of a Gaussian series. Hirota's theorem (6.3) gives ($w = -kc$)

$$\bar{\zeta}(n-n'c, A) = (2\pi i k)^4 (n-n')^4 + (2\pi i k)(2\pi i w)(n-n')^2 - 2A, \qquad (6.6)$$

which is a function only of the *difference* $n-n'$. As shown in Sec. 4, this property, that ζ (or $\bar{\zeta}$) is a function only of $n-n'$, is sufficient to prove that the vanishing of two (possibly nonlinear) equations in c and A is sufficient to give ρ (or $\bar{\rho}$)$\equiv 0$. Nakamura has shown that this generalizes to higher-dimensional theta functions, too—an N-dimensional Fourier series yields 2^N nonlinear equations in the phase speeds, A and the parameters of the theta function series that are sufficient to determine the whole solution. Furthermore, because the fact that $\bar{\zeta}$ is a function only of the difference $n-n'$ is a direct consequence of the fact that the bilinear operators yield results that depend only on $(k-k')$ and $(w-w')$ as shown in (6.3), it follows that the theta function solutions to *all* of Hirota's transformed equations expressed in terms of D_x and D_t must also be reducible to 2^N nonlinear algebraic equations in the parameters.

Our goal in this section is to prove a generalization of (6.3) which is applicable to Gaussian series. We shall find that, just as for Fourier series, the action of D_t and D_x on a pair of Gaussians depends only on the difference in their arguments. This in turn immediately implies that when an N-dimensional Gaussian theta function is substituted into any of Hirota's equations, the problem again reduces to 2^N nonlinear algebraic equations.

Before stating and proving the central result, it is important to note one powerful simplication: all terms in a given theta–Gaussian series have *identical second degree arguments* in the exponentials, and thus differ only in the terms in the exponentials which are *linear* in x and t. This is because

all the terms in the theta series are Gaussians of *identical* shape, differing only in the location of the peak. Consequently, one can always factor out the common second-degree exponential as was done explicitly for the one-dimensional theta function in (2.9). The same will be done in the results below.

Theorem: Let D_x, D_t be the usual Hirota bilinear operators defined by (6.1) above, and let

$$F \equiv \exp[-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2]\exp(kx + wt), \tag{6.7}$$

$$G \equiv \exp[-(\alpha/2)x^2 - \beta xt - (\gamma/2)t^2]\exp(k'x + w't); \tag{6.8}$$

then

(i)

$$D_x^n(F\cdot G) = \alpha^{n/2} H_n\left[\frac{k-k'}{2\alpha^{1/2}}\right]FG, \tag{6.9}$$

where $H_n(y)$ is the usual Hermite polynomial. $D_t^n(F\cdot G)$ is given by (6.9) also with α replaced by γ and $(k-k')$ by $(w-w')$.

(ii)

$$\exp[\delta D_x](F\cdot G) = \exp[-\alpha\delta^2 + \delta(k-k')]FG. \tag{6.10}$$

(iii) Defining

$$Q_n^m \equiv D_x^n D_t^m (F\cdot G), \tag{6.11}$$

Q_n^m is determined by the recursion

$$Q_n^{m+1} = -2n\beta Q_{n-1}^m + (w-w')Q_n^m - 2\gamma m Q_n^{m-1}, \tag{6.12}$$

where the starting values are

$$Q_n^{-1} \equiv 0 \text{ for all } n, \tag{6.13}$$

$$Q_n^0 = \alpha^{n/2} H_n\left[\frac{k-k'}{2\alpha^{1/2}}\right] \quad \text{[as given by (i)]}. \tag{6.14}$$

Proof: The demonstration of (i) is inductive. We use Q_n^m as defined in (6.11) except that we do *not* set $x' = x$, $t' = t$ until after obtaining the general recursion. We also temporarily drop the superscript, which is understood to be 0.

Define

$$Q_0 \equiv FG. \tag{6.15}$$

By explicit differentiation

$$Q_1 = D_x(F\cdot G) = [-\alpha(x-x') - \beta(t-t') + (k-k')]FG \tag{6.16}$$

To proceed to the next order, we can use the Leibnitz product-of-a-derivative rule after replacing FG by Q_0 to obtain

$$\begin{aligned} Q_2 &= D_x^2(F\cdot G) \\ &= [-\alpha(x-x') - \beta(t-t') + (k-k')]D_x(Q_0) \\ &\quad + Q_0 D_x[-\alpha(x-x') - \beta(t-t') + (k-k')] \end{aligned} \tag{6.17}$$

$$= [-\alpha(x-x') - \beta(t-t') + (k-k')]Q_1 - 2\alpha Q_0. \tag{6.18}$$

Let us now suppose that the recursion relation

$$Q_{n+1} = [-\alpha(x-x') - \beta(t-t') + (k-k')]Q_n - 2n\alpha Q_{n-1} \tag{6.19}$$

holds for a given n; we have already shown that it is true for $n = 1$. To demonstrate that it must also hold with $n \to n+1$, again apply D_x to (6.19) and invoke the Leibnitz rule to obtain

$$\begin{aligned} Q_{n+2} &= Q_n D_x[-\alpha(x-x') - \beta(t-t') + (k-k')] \\ &\quad + [-\alpha(x-x') - \beta(t-t') + (k-k')]D_x Q_n \\ &\quad - 2n\alpha D_x Q_{n-1} \end{aligned} \tag{6.20}$$

$$\begin{aligned} &= -2\alpha Q_n + [-\alpha(x-x') - \beta(t-t') \\ &\quad + (k-k')]Q_{n+1} - 2n\alpha Q_n, \end{aligned} \tag{6.21}$$

which is identical with (6.19) except for the replacement of n by $n+1$. Therefore, by induction, the recursion (6.19) must hold for all n. Setting $x = x'$ and $t = t'$ simplifies it to

$$Q_{n+1} = (k-k')Q_n - 2n\alpha Q_{n-1}. \tag{6.22}$$

The Hermite polynomials $H_n(y)$ satisfy the recursion relation

$$H_{n+1} = 2yH_n - 2nH_{n-1}. \tag{6.23}$$

It is trivial to show that

$$Q_n = \alpha^{n/2} H_n\left[\frac{k-k'}{2\alpha^{1/2}}\right] \tag{6.24}$$

satisfies (6.22) and the starting values (6.15) and (6.16) by directly substituting (6.24) into (6.22) and then employing the Hermite recursion (6.23) with the starting values $H_0 = 1$ and $H_1 = 2y$.

Part (ii) is derived by interpreting the exponential of an operator as the power series

$$\exp[\delta D_x] \equiv \sum_{n=0}^{\infty} \frac{\delta^n D_x^n}{n!}, \tag{6.25}$$

applying (i) term by term, and then equating the result to the generating function of the Hermite polynomials

$$e^{-t^2+2ty} = \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(y), \tag{6.26}$$

with $t = \delta\alpha^{1/2}$ and $y = (k-k')/2\alpha^{1/2}$.

An alternative proof can be obtained by using Hirota's result[4] that $\exp(\delta D_x)(F\cdot G) = F(x+\delta)G(x-\delta)$ for any F, G, and specializing it to the case when F and G are both Gaussians.

Part (iii) is proved by induction and use of the Liebnitz rule exactly as for (i) with the addition of the identity of $dH_n/dy = 2nH_{n-1}$, so details will be omitted. A simple, closed form solution for (6.12) has not been found; however, since Hirota's various evolution equations involve only first or second mixed derivatives, such a general solution is not really needed for the theory of polycnoidal waves.

7. MULTIDIMENSIONAL THETA FUNCTIONS

The general Riemann theta function of "reduced half integer characteristic" and dimension N is defined by

$$\begin{aligned} &\theta\begin{bmatrix}\boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}'\end{bmatrix}(\delta, \mathbf{T}) \\ &= \sum_{\mathbf{n}} \exp\left\{\pi i\left[\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}\left(n_i + \frac{\epsilon_i}{2}\right)\right.\right. \\ &\quad \left.\left.\times\left(n_j + \frac{\epsilon_j}{2}\right) + 2\sum_{i=1}^{N}\left(n_i + \frac{\epsilon_i}{2}\right)\left(\zeta_i + \frac{\epsilon_i'}{2}\right)\right]\right\}. \end{aligned} \tag{7.1}$$

$\mathbf{T}$ is the $N \times N$ symmetric square matrix, called the "theta matrix," with positive definite *imaginary* part whose ele-

ments are written T_{ij}. ξ is the N-dimensional vector of dependent variables; in applications to polycnoidal waves

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{pmatrix} = \begin{pmatrix} k_1(x - c_1 t) + \phi_1 \\ k_2(x - c_2 t) + \phi_2 \\ \vdots \\ k_N(x - c_N t) + \phi_N \end{pmatrix}, \tag{7.2}$$

where the k_i are wave numbers, the c_i phase speeds, and the ϕ_i are constant phase factors. The quantity $\begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix}$ consists of two N-dimensional row vectors written one above the other, where each element is either 1 or 0 and is known as the "characteristic" of the theta function,

$$\begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix} = \begin{pmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_N \\ \epsilon_1' & \epsilon_2' & \cdots & \epsilon_N' \end{pmatrix}. \tag{7.3}$$

Since there are a total of $2N$ matrix elements in the characteristic, each of which can independently take 2 values, there are a total of 4^N linearly independent theta functions with reduced half-integer characteristics. Note that it is conventional to define the multidimensional theta function so that it is periodic with period 2 versus the period of π that is conventional for the one-dimensional Jacobian theta functions θ_3 and θ_4.

Fortunately, it is always sufficient to take $\Theta \begin{bmatrix} 0 \\ 0 \end{bmatrix}(\xi; \mathbf{T})$ as the solution of any of Hirota's differential equations, where $\mathbf{0}$ is the N-dimensional vector whose elements are 0. Hirota and Ito[15] have shown that when the Fourier series of $\Theta \begin{bmatrix} 0 \\ 0 \end{bmatrix}(\xi, \mathbf{T})$ is substituted into one of the evolution equations, the result is a residual of the form

$$\bar{\rho} = \sum_{\epsilon_1 = 0}^{1} \sum_{\epsilon_2 = 0}^{1} \cdots \sum_{\epsilon_N = 0}^{1} R(\epsilon) \theta \begin{bmatrix} \epsilon \\ \mathbf{0} \end{bmatrix}(2\xi, 2\mathbf{T}). \tag{7.4}$$

There are a total of 2^N reduced half-integer theta functions with $\epsilon' \equiv 0$, and thus there are 2^N terms in (7.4) and 2^N nonlinear equations

$$R(\epsilon_1, \epsilon_2, \ldots, \epsilon_N) = 0, \quad \epsilon_i = 0 \text{ or } 1 \tag{7.5}$$

that must be solved to obtain the N phase speeds c_N, the constant of integration A in the H–KdV equation, and the $N(N-1)/2$ off-diagonal theta elements. (Recall $T_{ij} = T_{ji}$.) For $N \geq 3$, this gives more equations than unknowns; Hirota and Ito[15] have shown numerically that for $N = 3$, one of the eight equations (7.5) is redundant, and that one must solve seven equations in seven unknowns. Presumably something similar happens for large N although an analytical proof is lacking.

The diagonal elements play a role analogous to that of the nome for ordinary elliptic functions. In one dimension

$$q = e^{\pi i T_{11}}, \tag{7.6}$$

where T_{11} is positive imaginary. In more dimensions, one can define "nomes" via

$$q_j = e^{\pi i T_{jj}}, \quad j = 1, \ldots, N \tag{7.7}$$

and obtain perturbative solutions in the form of an N-dimensional power series in the q_j.

The "theta constants" are defined by

$$\theta \begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix}(T) \equiv \theta \begin{bmatrix} \epsilon \\ \epsilon' \end{bmatrix}(\mathbf{0}, \mathbf{T}). \tag{7.8}$$

Hirota and Ito,[15] without calling attention to the fact, show that the residual equations (7.5) can be expressed entirely in terms of theta functions of the wave numbers $k_1, \ldots, k_N$ [when the Hirota equation involves $\exp(\delta D_x)$] and the theta constants $\theta \begin{bmatrix} \epsilon \\ \mathbf{0} \end{bmatrix}(2\mathbf{T})$ and their derivatives with respect to the *diagonal* matrix elements. This is analogous to Sec. 4 where the solution for the cnoidal was expressed in terms of the functions $H(q')$ and $I(q')$—which effectively *are* theta constants—and their first two derivatives with respect to q'. In general, if the highest power of $D_x^n D_t^m$ is $m + n = J$, the residuals $R(\epsilon)$ will involve differentiations with respect to T_{jj} of up to order $J/2$; note that only *even* values of $(m + n)$ occur in Hirota's equations.

The significance of this (besides the fact that it provides a simple and compact description of the residuals) is that it implies that one can apply Poisson summation directly to the residuals $R(\epsilon)$ to transform them into nonlinear equations in the complementary nomes q'_1, q'_2, etc.

The only flaw with this is that "theta-matrix doubling" occurs; the theta constants will appear in the residual equations $R(\epsilon)$ have double the theta matrix of the theta function that solves the H–KdV equation. In terms of the nomes, this means that the perturbation series for $c_1, \ldots, c_N$ and so on are functions of the *squares* of $q_1, q_2, \ldots, q_N$ rather than the nomes themselves. This implies very rapid convergence, of course, and is the reason that the first and second approximations to c in the one-dimensional case were seen in Sec. 5 to give such remarkable accuracy.

However, the Poisson summation for the theta function is given by

$$\theta \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}(\xi, \mathbf{T}) = \frac{\pi^{N/2}}{|\det T|^{1/2}} \theta \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}(\mathbf{T}^{-1}\xi, \mathbf{T}^{-1}), \tag{7.9}$$

which shows that doubling the theta matrix halves that of the Poisson sum. In other words, by applying Poisson summation directly to the residual equations $R(\epsilon)$, we pay for the rapidly-converging series in q^2 by obtaining slowly-convergent series in $(q')^{1/2}$ upon Poisson summation. Equations (4.20) and (4.21) show that exactly the same happens in reverse when we attempt to write theta constant series in q' directly in terms of those in q.

Thus, the best approach is to substitute *separately* the Fourier series and the Gaussian series into the H–KdV equations and its fellows, to obtain a series for c in q^2 from the Fourier series and another in q'^2 from the Gaussian series.

When performing this direct substitution for the Gaussian series, it is convenient to use the freedom to shift the phase factors in $\xi_1, \xi_2, \ldots, \xi_N$ to replace $\theta \begin{bmatrix} 0 \\ 0 \end{bmatrix}(\xi, \mathbf{T})$ by $\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix}(\xi, \mathbf{T}) = \theta \begin{bmatrix} 0 \\ 0 \end{bmatrix}(\xi + \frac{1}{2}, \mathbf{T})$ because this has the simpler Poisson sum

$$\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix}(\xi, \mathbf{T}) = \frac{\pi^{N/2}}{|\det T|^{1/2}} \times \sum_n \exp\left\{ -\pi \sum_{i=1}^{N} \sum_{j=1}^{N} (\xi_i + n_i)(\xi_j + n_j) S_{ij} \right\}, \tag{7.10}$$

where the sum is taken over all possible half integers $\{ -\infty, \ldots, -\frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots, \infty \}$ in each of the N sum variables

$n_1, \ldots, n_N$, and where

$$S = I_m(\mathbf{T}^{-1}). \tag{7.11}$$

In one dimension, $\theta\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right](\xi, \mathbf{T}) = \theta_4(x; q)$, which was what we used in Secs. 2–5.

Again, however, the substitution of the series reduces the problem—after using the theorems of Sec. 6—to solving 2^N nonlinear equations in $N^2/2 + N/2 + 1$ unknowns, implying some redundancy in the residual equations, just as when the theta Fourier series was used. Again, these equations can be solved perturbatively in the N complementary nome variables defined by

$$q_j' = e^{-\pi s_{jj}}. \tag{7.12}$$

The lowest 2^N terms in (7.10), those with $n_j = \pm 1/2$ for all j, generate the N-soliton solution of the KdV equation just as the bi-Gaussian ($N = 1$) was shown to generate the single soliton in Sec. 3. For $N = 2$, we have a "tetra-Gaussian" whose four peaks form a rectangle in the ζ_1-ζ_2 plane; for $N = 3$, an "octo-Gaussian" whose eight centers form the corners of a cube in ζ_1–ζ_2–ζ_3 space. And so it goes.

We close with two elementary theorems which simplify the calculations and have already been used above.

Theorem:

(i) If $U(\zeta_1, \zeta_2, \ldots, \zeta_N)$ is a solution of the KdV equation where $\zeta_j = k_j(x - c_j t) + \phi_j$ as in (7.2), then

$$V(x, t) = p + U(\bar{\zeta}_1, \bar{\zeta}_2, \ldots, \bar{\zeta}_N) \tag{7.13}$$

is also a solution of the KdV equation provided

$$\bar{\zeta}_j(k_j, \phi_j, c_j) = \zeta_j(k_j, \phi_j, c_j + p) \tag{7.14}$$

for all j where p is an arbitrary constant.

(ii) If $u(x, t)$ is a solution of the KdV equation, then

$$\tilde{v}(x, t) \equiv \lambda^2 u(\lambda x, \lambda^3 t) \tag{7.15}$$

is also a solution.

The proofs are elementary and are not given here. The first theorem allows one to choose "mean sea level," i.e., $\int_{-\pi/2}^{\pi/2} u(x, t)\, dx$, to be whatever one wishes. Note that the theta-function solution normally picks its own "mean sea level," which is generally different from zero.

The second theorem allows us to take the periodicity interval to be π or 2 or whatever is convenient. It also permits us to set one of the wave numbers $k_1 = 1$ without loss of generality.

8. CONCLUSIONS AND SUMMARY

The theory of polycnoidal waves for the Korteweg–de Vries and other evolution equations is built upon four fundamental ideas. The first was the recognition by Lax and Novikov that the KdV equation had a class of generalized cnoidal wave solutions, here dubbed polycnoidal waves, that could be used to approximate an arbitrary, spatially-periodic initial condition and that could be formally calculated from the spectrum of Hill's equation. The second was the independent discovery by Akira Nakamura and the author that Hirota's "direct method" was just as useful for polycnoidal waves as for the multiple-soliton solutions of which they are generalizations. Nakamura, Hirota, and Ito subsequently refined the "direct method" using Fourier series to a high art.

The third is the discovery, first presented here, that the direct method can also be applied using the alternative Gaussian series for the theta functions to make it possible to explore strongly nonlinear polycnoidal waves and their relationship with multiple solitary waves.

The fourth, also presented here for the first time, is that by using these two different series—the theta Fourier expansion and the theta Gaussian expansion—in a complementary way, the former for small amplitude and the latter for large amplitude, one can obtain rapidly convergent perturbation series—a couple of terms are enough—to approximate the polycnoidal waves over the whole range of parameter space.

There are some additional complications, such as the way soliton phase shifts enter the formalism, which arise for polycnoidal waves of dimension $N \geq 2$. Since this present work is already lengthy, the actual polycnoidal wave calculations will be described in a later work. Here, however, the full mathematical machinery to perform these calculations has been presented with a very thorough discussion of the ordinary cnoidal wave to illustrate both how to use the formalism and also why perturbation theory should be useful even for the more complicated cases. As we saw in Sec. 5, just two terms of the perturbation series in q (derived from the theta Fourier series) and two terms of the series in q' (derived from the theta Gaussian series) were sufficient to give the cnoidal wave phase speed c to within a relative error of 0.027% for all possible values of q and q'.

The path for future research is to simply follow up this initial success by explicit calculations for $N \geq 2$, concentrating particularly on the Gaussian series approach that is the central theme of this work. For the Fourier series results of Nakamura and Hirota can be obtained another way as explained in Appendix B—an even more direct method than Hirota's. However, at present there seems to be no alternative to the theta Gaussian series for exploring strongly nonlinear polycnoidal waves and their close relationship with multiple solitons.

ACKNOWLEDGMENT

This work was supported by NSF Grant OCE-79-09191.

APPENDIX A: THE EXACT SOLUTION FOR THE CNOIDAL WAVE

By using identities 17.2.13 and 17.2.11 of the NBS Handbook,[23] one can show that

$$12[\ln \theta_4]_{XX} = \delta + \epsilon cn^2(2K(m)X/\pi; m), \tag{A1}$$

where $K(m)$ is the complete elliptic integral with m related to q by

$$q = e^{-\pi K(1-m)/K(m)} \tag{A2}$$

and with the phase speed c given by ($X = x - ct$)

$$c = \delta + [(2m - 1)/3m]\epsilon, \tag{A3}$$

with

$$\delta = [-48K(m)/\pi^2][(m - 1)K(m) + E(m)], \tag{A4}$$

$$\epsilon = -48mK^2(m)/\pi^2. \tag{A5}$$

The phase speed is obtained by substituting the right-hand side of (A1) directly into the KdV equation and using elliptic

function identities.[23] For small q, one has the approximations

$$m = 16q(1 - 8q + 44q^2), \tag{A6}$$

$$K(m(q)) \doteq (\pi/2)(1 + 4q + 4q^2), \tag{A7}$$

$$E(m(q)) \doteq (\pi/2)(1 - 4q + 20q^2). \tag{A8}$$

Substituting these small q approximations into (A4) and (A5) gives the approximation for c given by (5.9). The corresponding approximation for large q is derived directly by solving the residual equations in Sec. 4, and then Taylor-expanding (4.28), which is actually a rational function of q'.

When carrying out the analysis in terms of theta functions, it is convenient to take q or q' as the perturbation parameter. In analyzing observations or laboratory experiments, one would probably take ϵ as the fundamental quantity since this is what is most easily measured. (It is the difference between the peak and the trough.) In applying multiple-scales perturbation theory to the KdV equation as in Appendix B, one would normally fix the Fourier coefficient of $\cos(2nx)$ (the linear wave) at a certain value a and use that as the perturbation parameter.

Fortunately, it is easy to relate these different measures of the nonlinearity of the wave to each other by using the Fourier series for $u(x, t) = 12(\ln \theta_4)_{xx}$ given earlier:

$$u(x, t) = 96q \sum_{n=1}^{\infty} \frac{nq^{n-1}}{1 - q^{2n}} \cos(2nx). \tag{A9}$$

The coefficient of $\cos(2x)$ is

$$a = 96q/(1 - q^2), \tag{A10}$$

which is trivially solved to give q as a function of a. Since $\epsilon = u(0, 0) - u(\pi/2, 0)$, we can evaluate (A9) at these values of x and subtract to obtain

$$\epsilon = 192q \sum_{n=0}^{\infty} \frac{(2n + 1)q^{2n}}{1 - q^{2(2n+1)}}, \tag{A11}$$

which can be easily reverted term-by-term to give a series expansion for $q(\epsilon)$.

APPENDIX B: THE METHOD OF MULTIPLE SCALES AND POLYCNOIDAL WAVES

In the near linear regime, one can bypass the use of the Fourier series for the theta functions by using a much more general technique[16,17] known variously as the "method of strained parameters" or "method of multiple scales," which can be applied directly to Korteweg–de Vries equation. The amplitude of the lowest harmonic, a, is assumed to be a small parameter. It is further assumed that (i) the wave is steadily translating at a phase speed c and (ii) that $u(x - ct)$ can be expanded as a power series in a. One can then substitute the power series into the differential equation, match powers of a and solve the perturbation equations order-by-order. However, there is one modest complication: the phase speed c is usually altered by the nonlinearity, so it is necessary to assume c can also be expanded in a power series in a. The technique is very similar to the usual Rayleigh–Schrödinger perturbation theory of quantum mechanics with c playing the role of the eigenvalue. A full description with many, many examples and problems is given in the texts by Nayfeh[16] and Bender and Orszag.[17] It is not exactly a new idea; Stokes applied it to water waves in 1847.

Because this algorithm is so straightforward, it is not only easy to do by hand but also simple to program for a computer. Using the algebraic manipulation language REDUCE 2, which can explicitly multiply and differentiate Fourier series and manipulate trigonometric identities, a short program was written by the author to calculate single and double cnoidal waves. For the expenditure of $1.50 (about 8 sec of CPU time on the University of Michigan Amdahl), the following was obtained to fifth order for the single cnoidal wave ($X = x - ct$):

$$c = -4 + a^2/96 + a^4/884\,736, \tag{B1}$$

$$\begin{aligned} u(x, t) = {} & a\cos(2X) + (a^2/48 - a^4/221\,184)\cos(4X) \\ & + (a^3/3072 - a^5/9\,437\,184)\cos(6x) \\ & + (a^4/221\,184)\cos(8X) \\ & + (5a^5/84\,934\,656)\cos(10X). \end{aligned} \tag{B2}$$

By Taylor-expanding the relationship between a and the elliptic function nome q, which was shown in Appendix A to be

$$a = 96q/(1 - q^2) = 96q(1 + q^2 + q^4 + \cdots), \tag{B3}$$

one can recast the expansion in powers of q and thus duplicate the results of the "direct Fourier series" method of Nakamura, Hirota, and others:

$$c = -4 + 96q^2 + 288q^4 + \cdots, \tag{B4}$$

$$\begin{aligned} u(x, t) = {} & 96q[(1 + q^2 + q^4)\cos(2x) \\ & + 2q\cos(4x) + 3q^2\cos(6x) \\ & + 4q^3\cos(8x) + 5q^4\cos(10x)]. \end{aligned} \tag{B5}$$

In the body of this paper, only terms through q^2 were kept because these give more than enough accuracy; the expansions have been carried to higher order here simply to make the point that it is easy to calculate Fourier series via the method of multiple scales and that a theta function approach is not really needed except in the opposite near-soliton regime of strong nonlinearity.

In a similar way, one can calculate double cnoidal waves by starting with the lowest order approximation

$$u(x, t) \doteq a(\cos(2X) + b\cos(MY)), \tag{B6}$$

where again a is a small parameter and b is $O(1)$, with

$$X = x - c_1 t + \phi_1, \tag{B7}$$

$$Y = x - c_2 t + \phi_2. \tag{B8}$$

M may take on any constant value; however, unless $M/2$ is a rational number, the wave will be "almost periodic" in x in the formal mathematical sense of the word as opposed to truly periodic. But Novikov[10] and Dubrovin *et al.*[11] have emphasized that the polycnoidal wave may indeed be almost periodic in space.

The most reasonable value of M, however, is $M = 4$ so that the second component is the second harmonic of the first. Unless the initial condition is rather peculiar, the second harmonic is usually the largest Fourier component after the fundamental. Consequently, in applying the double cnoidal wave to model events in a laboratory tank or the real world, $M = 4$ is the case in which one would be chiefly inter-

ested. For simplicity, the results below are therefore confined to this case. The computer program, however, can solve the double cnoidal wave with M as a purely symbolic parameter with no numerical value specified.

One obtains

$$c_1 = -4 + a^2/96, \tag{B9}$$

$$c_2 = -16 + a^2 b^2/384, \tag{B10}$$

$$\begin{aligned} u(x,t) = {} & a[\cos(2X) + b\cos(4Y)] \\ & + a^2[(1/48)\cos(4X) + (b^2/192)\cos(8Y) \\ & + (b/48)\cos(2X+4Y) - (b/48)\cos(2X-4Y)] \\ & + a^3[(1/3072)\cos(6X) + (b^3/49152)\cos(12Y) \\ & + (b/2592)\cos(4X+4Y) \\ & + (25b^2/165888)\cos(2X+8Y) \\ & - (b^2/2048)\cos(2X-8Y)]. \end{aligned} \tag{B11}$$

This run also cost a mere \$1.50. It is trivial to generalize the algorithm to triple and higher cnoidal waves; in point of fact, the same program with changes in only two statements was used to compute both the single and double cnoidal wave results given above.

The message is clear. The theta Fourier series is not essential in understanding spatially periodic solutions of the KdV and other evolution equations; it is the theta Gaussian series that contains treasure.

[1]D. J. Korteweg and G. de Vries, Philos. Mag. **39**, 422 (1895).
[2]J. Scott-Russell, Proc. Roy. Soc. Edinburgh **319** (1844).
[3]C. S. Gardner, J. M Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[4]R. Hirota, Prog. Theor. Phys. **52**, 1498 (1974).
[5]Ryogo Hirota, Phys. Rev. Lett. **27**, 1192 (1971).
[6]J. P. Boyd, J. Phys. Oceangr. **10**, 1699 (1980).
[7]Y.-C. Ma and L. G. Redekopp, Phys. Fluids **22**, 1872 (1979).
[8]Y.-C. Ma, Stud. Appl. Math. 59, **201** (1978).
[9]P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975); H. P. McKean and P. van Moerbeke, Inv. Math **30**, 217 (1975).
[10]S. P. Novikov, Funct. Anal. Appl. **8**, 236 (1974).
[11]B. A. Dubrovin, V. B. Matreev, and S. P. Novikov, Russian Math. Surveys **31**, 59 (1976).
[12]Akira Nakamura, J. Phys. Soc. Japan **47**, 1701 (1979).
[13]Akira Nakamura, J. Phys. Soc. Japan **48**, 1365 (1980).
[14]Akira Nakamura and Yoshimasa Matsuno, J. Phys. Soc. Japan **48**, 653 (1980).
[15]Ryogo Hirota and Masaaki Ito, J. Phys. Soc. Japan **50**, 338 (1981).
[16]A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).
[17]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[18]Richard Bellman, *A Brief Introduction to Theta Functions* (Holt, Rinehart, and Winston, New York, 1961), p. 45.
[19]The Fourier series of a singular function which is analytic on the real axis has a finite strip of convergence in the complex plane $\{\text{all } x;\ -\tau < y < \tau\}$, where $z = x + iy$ and the singularity of $f(z)$ nearest the real axis lies on $y = \pm\tau$. Since $|\cos(nz)|$ is bounded by $\exp(ny)$, it follows that the Fourier series can converge inside the strip bounded by the lines on which $f(z)$ is singular and diverge outside it only if the Fourier coefficients a_n satisfy the inequality $\exp[-n(\tau+\epsilon)] < |a_n| < \exp[-n(\tau-\epsilon)]$. Subtracting off the poles on $y = \pm\tau$, using the known Fourier expansion for $1/(a+\cos z)$ and noting that $f(z) - 1/(a+\cos z)$ has a larger strip of convergence than $f(z)$ can be used to establish that $a_n \sim O(q^n)$ precisely, where $q = \exp(-\tau)$.
[20]Richard Bellman, Ref. 18, pp. 7–12; Philip M. Morse and Herman Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), pp. 466–467.
[21]Strictly speaking, $\Theta(X)$ must be multiplied by an X-independent constant to write it as the sum of two Gaussians. However, a constant times a solution of the H-KdV equation is also a solution, so this multiplicative constant has no physical significance.
[22]This is a corollary of a theorem which will be proved in Sec. 6 although it has been known to soliton specialists for years.
[23]Milton Abramowitz and Irene Stegun, Eds., *Handbook of Mathematical Functions* (Dover, New York, 1965).
[24]For notational simplicity, θ_3 is used here instead of θ_4, and the Fourier sum must be taken over the integers instead of the half-integers. These minor differences are irrelevant to the analogy between (4.2) and (6.5), however.

String mechanics based on 2-forms

P. Mitra
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay-400 005, India

(Received 22 October 1980; accepted for publication 23 December 1980)

Nambu invented a mechanics for strings by replacing the fundamental 1-form $p_i dq^i - H\,dt$ of Hamiltonian mechanics by a certain 2-form. We study the mechanics corresponding to a more general 2-form applicable to weighted strings. Our equations of motion are fully deterministic, unlike those of Nambu, which need a supplementary condition. We set up a Hamilton–Jacobi formalism closely paralleling ordinary mechanics.

PACS numbers: 03.40. – t

1. INTRODUCTION

Classical Hamiltonian mechanics can be said to be based on the differential 1-form[1] $p_i dq^i - H\,dt$. Here q^i $(i = 1,2,...,n)$ are the generalized coordinates of a dynamical system, p_i the conjugate momenta, t is the time, and $H(q,p)$ the Hamiltonian function. Nambu[2] recently suggested the use of a 2-form as the starting point, with the single independent variable t replaced by two independent variables σ and τ. He was interested in the dynamics of strings, where one needs two parameters to describe the sheet traced out in spacetime by a moving string. For constructing a 2-form he took the wedge products $dq^i \wedge dq^j$ and $d\sigma \wedge d\tau$. Introducing "momenta" $\pi_{ij} = -\pi_{ji}$ [only $n(n-1)/2$ are independent] and a Hamiltonian function $H(q,\pi)$, he wrote down the 2-form $\frac{1}{2}\pi_{ij} dq^i \wedge dq^j - H\,d\sigma \wedge d\tau$ and used it to derive equations of motion. He also suggested that the analog of the Hamilton–Jacobi equation

$$dS = p_i dq^i - H\,dt \tag{1}$$

should be

$$\sum_{m=1}^{2} dS_m \wedge dT_m = \tfrac{1}{2}\pi_{ij} dq^i \wedge dq^j - H\,d\sigma \wedge d\tau, \tag{2}$$

where the manifold considered is, of course, the extended configuration space (coordinates: q^i,σ,τ) rather than the extended phase space (ccordinates: q^i,π_{ij},σ,τ).

While Nambu's mechanics may be appropriate for some string systems, its governance of time evolution is not quite as in ordinary Hamiltonian mechanics. In ordinary mechanics, the equations of motion are sufficient for the purpose of determining the second- and higher-order time derivatives of the coordinates q^i if the initial values of q^i and their first-order time derivatives are given. In Nambu's mechanics, however, the equations of motion have to be supplemented by a relation connecting the time τ with the coordinates q^i before this can be done.[3]

In this paper, we study the mechanics generated by a differential 2-form which includes terms of the type $d\sigma \wedge dq^i$ and $dq^i \wedge d\tau$ in addition to those considered by Nambu. This 2-form, besides being very general, merits consideration for a practical reason too: it is expected to be relevant for weighted strings, according to Nambu himself.[2] Furthermore, the resulting equations of motion are fully deterministic, i.e., do not share with Nambu's equations of motion the peculiarity referred to above.

After discussing the equations of motion and the initial-value problem in Sec. 2, we take on the Hamilton–Jacobi (HJ) equation in Sec. 3. A special aim of this paper is to establish an HJ formalism parallel to the one of ordinary mechanics. Unlike Eq. (2), our HJ system involves only one unknown function(al). We show that our systems of equations is satisfied by the action functional and explain how a complete solution can be used to integrate the equations of motion by a transformation to constant coordinates and momenta.

2. EQUATIONS OF MOTION FROM 2-FORMS

In Hamiltonian mechanics, the 1-form

$$\omega^1 = p_i dq^i - H\,dt, \tag{3}$$

defined on the extended phase space (coordinates: q^i, p_i, t), gives rise to the 2-form

$$d\omega^1 = dp_i \wedge dq^i - \frac{\partial H}{\partial q^i} dq^i \wedge dt - \frac{\partial H}{\partial p_i} dp_i \wedge dt \tag{4}$$

by exterior differentiation. If $q^i = q^i(t)$, $p_i = p_i(t)$ are to be the equations describing a possible evolution curve of the system, one requires that the vector

$$X = \frac{dq^i}{dt}\frac{\partial}{\partial q^i} + \frac{dp_i}{dt}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t} \tag{5}$$

satisfy, at each point (q,p,t) on the curve, the equation

$$d\omega^1(X,Y) = 0 \tag{6}$$

for an arbitrary vector Y at that point. This immediately leads to Hamilton's equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q^i}. \tag{7}$$

We wish to give an analogous derivation of the equations of motion for strings. As announced in Sec. 1, we shall start from a 2-form containing terms of the kind $d\sigma \wedge dq^i$, $dq^i \wedge d\tau$, $dq^i \wedge dq^j$, and $d\sigma \wedge d\tau$. We introduce three sets of "momenta" p_i,ϕ_i,π_{ij} (antisymmetric in i and j as in Ref. 2) and a Hamiltonian function $H(q,p,\phi,\pi)$. In the extended phase space with local coordinates q^i, p_i,ϕ_i,π_{ij} (independent ones only), σ, and τ, our 2-form is

$$\omega^2 = p_i d\sigma \wedge dq^i + \phi_i dq^i \wedge d\tau + \tfrac{1}{2}\pi_{ij} dq^i \wedge dq^j - H d\sigma \wedge d\tau. \tag{8}$$

 0022-2488/82/030388-04$02.50 © 1982 American Institute of Physics

This gives rise to the 3-form[4]

$$d\omega^2 = dp_i \wedge d\sigma \wedge dq^i + d\phi_i \wedge dq^i \wedge d\tau + \tfrac{1}{2} d\pi_{ij} \wedge dq^i \wedge dq^j - \frac{\partial H}{\partial q^i} dq^i \wedge d\sigma \wedge d\tau - \frac{\partial H}{\partial p_i} dp_i \wedge d\sigma \wedge d\tau - \frac{\partial H}{\partial \phi_i} d\phi_i \wedge d\sigma \wedge d\tau - \tfrac{1}{2} \frac{\partial H}{\partial \pi_{ij}} d\pi_{ij} \wedge d\sigma \wedge d\tau. \quad (9)$$

If $q^i = q^i(\sigma,\tau)$, $p_i = p_i(\sigma,\tau)$, $\phi_i = \phi_i(\sigma,\tau)$, and $\pi_{ij} = \pi_{ij}(\sigma,\tau)$ are to be the equations describing a possible evolution surface of the string, we require that the vectors

$$X = \frac{\partial q^i}{\partial \sigma}\frac{\partial}{\partial q^i} + \frac{\partial p_i}{\partial \sigma}\frac{\partial}{\partial p_i} + \frac{\partial \phi_i}{\partial \sigma}\frac{\partial}{\partial \phi_i} + \frac{1}{2}\frac{\partial \pi_{ij}}{\partial \sigma}\frac{\partial}{\partial \pi_{ij}} + \frac{\partial}{\partial \sigma}, \quad (10)$$

$$Y = \frac{\partial q^i}{\partial \tau}\frac{\partial}{\partial q^i} + \frac{\partial p_i}{\partial \tau}\frac{\partial}{\partial p_i} + \frac{\partial \phi_i}{\partial \tau}\frac{\partial}{\partial \phi_i} + \frac{1}{2}\frac{\partial \pi_{ij}}{\partial \tau}\frac{\partial}{\partial \pi_{ij}} + \frac{\partial}{\partial \tau}, \quad (11)$$

satisfy at each point $(q, p, \phi, \pi, \sigma, \tau)$ of the surface, the equation

$$d\omega^2(X,Y,Z) = 0, \quad (12)$$

for an arbitrary vector Z at that point. The immediate consequence is the set of equations

$$\frac{\partial q^i}{\partial \sigma}\frac{\partial q^j}{\partial \tau} - \frac{\partial q^i}{\partial \tau}\frac{\partial q^j}{\partial \sigma} = \frac{\partial H}{\partial \pi_{ij}}, \quad (13)$$

$$\frac{\partial q^i}{\partial \sigma} = \frac{\partial H}{\partial \phi_i}, \quad (14)$$

$$\frac{\partial q^i}{\partial \tau} = \frac{\partial H}{\partial p_i}, \quad (15)$$

$$\frac{\partial \pi_{ij}}{\partial \sigma}\frac{\partial q^j}{\partial \tau} - \frac{\partial \pi_{ij}}{\partial \tau}\frac{\partial q^j}{\partial \sigma} + \frac{\partial p_i}{\partial \tau} + \frac{\partial \phi_i}{\partial \sigma} = -\frac{\partial H}{\partial q_i}. \quad (16)$$

Equations (13)–(16) will determine the dynamics of the string. However, not all of them can be regarded as equations of motion. Thus, $\partial q^i/\partial \sigma$ and $\partial q^i/\partial \tau$ being determined by (14) and (15), (13) becomes a constraint:

$$\frac{\partial H}{\partial \phi_i}\frac{\partial H}{\partial p_j} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial \phi_j} = \frac{\partial H}{\partial \pi_{ij}}. \quad (17)$$

We can imagine these equations to be solved for π_{ij} in terms of q, p, and ϕ. Now suppose that the values of q, p, and ϕ at $\sigma = \sigma_0$ and $\tau = \tau_0$ are known. Do the Eqs. (14)–(16) determine the values for other σ and τ? Clearly not, for the σ and τ derivatives of p and ϕ cannot all be determined from (16). More initial data are needed. If the coordinates and momenta are known for all values of σ at $\tau = \tau_0$ (say), the equations can predict the behavior for all τ. Obviously, σ and τ may be interchanged here, and may even be replaced by any pair of independent functions of themselves; however, we shall choose to think of τ as the time and σ as a parameter labelling different points of the string, so that initial data consist of values for all σ at fixed τ. Note that because of (14) it is not permissible to specify values of q, p, and ϕ for all σ arbitrarily. These equations can in fact be solved for ϕ_i in terms of q, p, and $\partial q/\partial \sigma$, so that it is sufficient to specify the initial values of q and p. Therefore we are to regard (14) as well as (17) as constraint equations, with the help of which π and ϕ are to be eliminated from (15) and (16). While (15) determines $\partial q^i/\partial \tau$, the system of Eqs. (16) will, in general, determine $\partial p/\partial \tau$. These are genuine equations of motion.

What happens if we consider 2-forms less general than (8)? Nambu[2] left out the $d\sigma \wedge dq^i$ and $dq^i \wedge d\tau$ terms and consequently obtained Eqs. (13) and (16) only, the latter of course without the p and ϕ contributions. One would think that from n out of the $n(n-1)/2$ Eqs. (13), the n quantities $\partial q^i/\partial \tau$ could be determined, the rest being used to constrain $n(n-1)/2 - n$ of the π_{ij}; the τ derivatives of the remaining π_{ij} could then be found from (16). Unfortunately, this cannot be done. If in Eq. (13) we replace $\partial q^i/\partial \tau$ by $\partial q^i/\partial \sigma$, we see that the left-hand side vanishes, so that there cannot be a unique solution for $\partial q^i/\partial \tau$. A detailed study of (13) and (16) shows that they do not determine all $\partial q^i/\partial \tau$ and $\partial \pi_{ij}/\partial \tau$ in terms of q, π, and their σ-derivatives, or alternatively, all $\partial^2 q^i/\partial \tau^2$ in terms of q, $\partial q/\partial \tau$, and their σ-derivatives. It is necessary to introduce a relation involving the q^i and τ to supplement the equations of motion.

If we were to start from a 2-form containing only the $d\sigma \wedge dq^i$ term in addition to the piece involving the Hamiltonian, we would be led to Eqs. (15) and (16), the latter without its ϕ and π terms. But these are just Hamilton's equations of motion, with the Hamiltonian density rather than the Hamiltonian functional on the right-hand side. Thus, this is the simplest—almost trivial—generalization from the 1-form to a 2-form. It teaches us that the coefficient of the $d\sigma \wedge dq^i$ term is to be regarded as the "ordinary" momentum. This, of course, is in line with the general situation where we got equations of motion for q^i and p_i, while the other "momenta" ϕ_i and π_{ij} were determined in terms of these by our constraint equations. No wonder that by leaving out this vital term Nambu obtained an unusual kind of mechanics. Of course, one can take the $dq^i \wedge d\tau$ term instead of the $d\sigma \wedge dq^i$ one—then one would again obtain Hamilton's equations, but with σ rather than τ as the evolution parameter, and ϕ_i rather than p_i as the momenta. If in addition to one of these we wish to include the $dq^i \wedge dq^j$ term, Eqs. (13) would again be treated as constraint equations and solved for π_{ij}. Lastly, if we were to take both the $d\sigma \wedge dq^i$ and the $dq^i \wedge d\tau$ terms but leave out $dq^i \wedge dq^j$, Eqs. (14) would be constraint equations to be solved for ϕ_i.

3. HAMILTON–JACOBI FORMALISM

We shall now formulate a variational principle for our equations. This will help us in setting up the HJ formalism. Let there be some functions $q^i(\sigma,\tau)$, $p_i(\sigma,\tau)$, $\phi_i(\sigma,\tau)$, and $\pi_{ij}(\sigma,\tau)$ satisfying Eqs. (13)–(16), i.e., representing a possible evolution of the string. We restrict our attention to any region D of the σ-τ plane. If we give small variations δq^i, δp_i, $\delta \phi_i$, and $\delta \pi_{ij}$ to the functions q^i, p_i, ϕ_i, and π_{ij}, how does the action

$$S = \iint_D d\sigma \wedge d\tau \left(\pi_{ij} \frac{\partial q^i}{\partial \sigma}\frac{\partial q^j}{\partial \tau} + \phi_i \frac{\partial q^i}{\partial \sigma} + p_i \frac{\partial q^i}{\partial \tau} - H \right) \quad (18)$$

change? We find

$$\delta S = \iint_D d\sigma \wedge d\tau \Big[\delta\pi_{ij} \frac{\partial q^i}{\partial \sigma} \frac{\partial q^j}{\partial \tau} + \pi_{ij} \Big(\frac{\partial \delta q^i}{\partial \sigma} \frac{\partial q^j}{\partial \tau} + \frac{\partial q^i}{\partial \sigma} \frac{\partial \delta q^j}{\partial \tau} \Big) + \delta\phi_i \frac{\partial q^i}{\partial \sigma} + \phi_i \frac{\partial \delta q^i}{\partial \sigma} + \delta p_i \frac{\partial q^i}{\partial \tau} + p_i \frac{\partial \delta q^i}{\partial \tau}$$
$$- \frac{\partial H}{\partial q^i} \delta q^i - \frac{\partial H}{\partial p_i} \delta p_i - \frac{\partial H}{\partial \phi_i} \delta\phi_i - \frac{1}{2} \frac{\partial H}{\partial \pi_{ij}} \delta\pi_{ij} \Big]$$
$$= \iint_D d\sigma \wedge d\tau \Big\{ \frac{1}{2} \delta\pi_{ij} \Big(\frac{\partial q^i}{\partial \sigma} \frac{\partial q^j}{\partial \tau} - \frac{\partial q^i}{\partial \tau} \frac{\partial q^j}{\partial \sigma} - \frac{\partial H}{\partial \pi_{ij}} \Big) + \delta\phi_i \Big(\frac{\partial q^i}{\partial \sigma} - \frac{\partial H}{\partial \phi_i} \Big) + \delta p_i \Big(\frac{\partial q^i}{\partial \tau} - \frac{\partial H}{\partial p_i} \Big)$$
$$- \delta q^i \Big[\frac{\partial}{\partial \sigma} \Big(\pi_{ij} \frac{\partial q^j}{\partial \tau} \Big) - \frac{\partial}{\partial \tau} \Big(\pi_{ij} \frac{\partial q^j}{\partial \sigma} \Big) + \frac{\partial p_i}{\partial \tau} + \frac{\partial \phi_i}{\partial \sigma} + \frac{\partial H}{\partial q^i} \Big]$$
$$+ \frac{\partial}{\partial \sigma} \Big(\pi_{ij} \delta q^i \frac{\partial q^j}{\partial \tau} + \delta q^i \phi_i \Big) - \frac{\partial}{\partial \tau} \Big(\pi_{ij} \delta q^i \frac{\partial q^j}{\partial \sigma} - \delta q^i p_i \Big) \Big\} = \oint_{\partial D} \Big[\Big(\pi_{ij} \frac{\partial q^j}{\partial \sigma} - p_i \Big) \delta q^i d\sigma + \Big(\pi_{ij} \frac{\partial q^j}{\partial \tau} + \phi_i \Big) \delta q^i d\tau \Big], \tag{19}$$

where we have used Eqs. (13)–(16) and Stokes' theorem: ∂D is the boundary of D. The final result (19) can be taken as our variational principle. While we have derived it from the equations of constraint and of motion, clearly we can reverse our route and derive these equations from the statement that (19) holds for arbitrary variations δq^i, δp_i, $\delta\phi_i$, and $\delta\pi_{ij}$. A simpler statement of the variational principle is that S is stationary with respect to variations δq^i, δp_i, $\delta\phi_i$, and $\delta\pi_{ij}$ that are arbitrary except for the restriction that δq^i vanishes on the boundary ∂D of the region of integration.

In ordinary mechanics,[1] the HJ equation is derived as a property of the action integral regarded as a function of the initial and final times and configurations. In the present case, specifying the string configuration means giving q^i as a function of σ at some time τ. So we shall have a functional $S(\tau_0,\tau_1)[q_0,q_1]$, depending on two times τ_0 and τ_1 and on the corresponding configurations $q_0^i(\sigma)$ and $q_1^i(\sigma)$. This is just the integral defined in Eq. (18), the region D of integration being the portion of the σ-τ plane lying between $\tau = \tau_0$ and $\tau = \tau_1$; the functions $q^i(\sigma,\tau)$, $p_i(\sigma,\tau)$, $\phi_i(\sigma,\tau)$, and $\pi_{ij}(\sigma,\tau)$ used in the integration are those solutions of Eqs. (13)–(16) that satisfy the conditions $q^i(\sigma,\tau_0) = q_0^i(\sigma)$ and $q^i(\sigma,\tau_1) = q_1^i(\sigma)$ for all σ. Of course the normal way of specifying initial data would be to give the values of $q^i(\sigma,\tau_0)$ and $p_i(\sigma,\tau_0)$ for all σ. In the given situation, we may imagine ourselves trying different values of $p_i(\sigma,\tau_0)$ and adjusting them till the resulting values of $q^i(\sigma,\tau_1)$ match the preassigned functions $q_1^i(\sigma)$. For small $|\tau_1 - \tau_0|$, $p_i(\sigma,\tau_0)$ may be expected to be fixed uniquely by this procedure.

From (19), we see directly how S changes with q_0^i and q_1^i:

$$\frac{\delta S}{\delta q_0^i(\sigma)} = \Big(\pi_{ij} \frac{\partial q^j}{\partial \sigma} - p_i \Big)_{\tau = \tau_0}, \tag{20}$$

$$\frac{\delta S}{\delta q_1^i(\sigma)} = \Big(p_i - \pi_{ij} \frac{\partial q^j}{\partial \sigma} \Big)_{\tau = \tau_1}, \tag{21}$$

provided the variation at the extremities of the string can be ignored.[5] To see how S changes if τ_1 is altered to $\tau_1 + \delta\tau_1$, we regard the integral as being made up of two parts—one from τ_0 to τ_1, the other from τ_1 to $\tau_1 + \delta\tau_1$. Noting that at τ_1, q^i changes by $-(\partial q^i/\partial\tau)\delta\tau_1$, we can evaluate the contribution of the first part by using (19); the second part can be evaluated directly from the definition (18). We get

$$\delta S = - \int d\sigma \Big(\pi_{ij} \frac{\partial q^j}{\partial \sigma} - p_i \Big) \Big(- \frac{\partial q^i}{\partial \tau} d\tau_1 \Big)$$
$$+ \int d\sigma \delta\tau_1 \Big(\pi_{ij} \frac{\partial q^i}{\partial \sigma} \frac{\partial q^j}{\partial \tau} + \phi_i \frac{\partial q^i}{\partial \sigma} + p_i \frac{\partial q^i}{\partial \tau} - H \Big)$$
$$= \delta\tau_1 \int d\sigma \Big(\phi_i \frac{\partial q^i}{\partial \sigma} - H \Big), \tag{22}$$

so

$$\frac{\partial S}{\partial \tau_1} = \int d\sigma \Big(\phi_i \frac{\partial q^i}{\partial \sigma} - H \Big)_{\tau = \tau_1}, \tag{23}$$

and similarly

$$\frac{\partial \sigma}{\partial \tau_0} = \int d\sigma \Big(H - \phi_i \frac{\partial q^i}{\partial \sigma} \Big)_{\tau = \tau_0}. \tag{24}$$

Equation (23) is essentially the HJ equation. It closely resembles the usual

$$\frac{\partial S}{\partial t} = - H(q,p)_{p = \partial S/\partial q}. \tag{25}$$

The analog of the instruction to substitute $\partial S/\partial q$ for p is our Eq. (21). Of course, we have some σ integrations because we are considering a continuum system and there are some contributions from the spurious momenta ϕ_i and π_{ij}.

To fortify our claim that (23) is the appropriate generalization of the usual HJ equation, we shall show that it can be used to effect a change of variables that takes us to constant coordinates. First let us pose the HJ problem properly. We have derived the HJ equation as a property of the action functional $S(\tau_0,\tau_1)[q_0,q_1]$. Now we shall consider an unknown functional $T(\tau)[q]$ constrained to satisfy the equation

$$\frac{\partial T}{\partial \tau} = \int d\sigma \Big(\phi_i \frac{\partial q_i}{\partial \sigma} - H \Big), \tag{26}$$

where ϕ,π on the right-hand side are understood to be expressed in terms of q, p by the constraint Eq. (14) and (17), and p itself is eliminated through

$$p_i - \pi_{ij}\frac{\partial q^j}{\partial \sigma} = \frac{\delta T}{\delta q^i(\sigma)}. \tag{27}$$

Since the left-hand side of (27) involves p_i not only explicitly but also implicitly through π_{ij}, this elimination cannot be carried out explicitly.

As in the ordinary Jacobi method of integrating the equations of motion, suppose that we have found a complete solution[6] for T, i.e. one which is a functional of n arbitrary functions. We write this as $T(\tau)[q,q']$, $q'^i(\sigma)$ being the new functions. We reintroduce the momenta p through (27). Since π is understood to be expressed in terms of q, p, the left-hand side of this equation is a functional of q, p while the right-hand side is a (τ-dependent) functional of q,q'. We may suppose these equations to be solved for q' as τ-dependent functionals of q, p. We also introduce functions p'_i by

$$p'_i = -\delta T/\delta q'^i(\sigma), \tag{28}$$

where the right-hand side is again to be thought of as a τ-dependent functional of q, p. We claim that the functions $q'^i(\sigma)$, $p'_i(\sigma)$, which give an alternative description of the state of the string, are constant coordinates, i.e., the explicit and implicit τ dependences of these functionals of q, p cancel each other. To prove this, we make use of the variational principle. Take any possible evolution of the string as represented by functions $q^i(\sigma,\tau)$, $p_i(\sigma,\tau)$, $\phi_i(\sigma,\tau)$, and $\pi_{ij}(\sigma,\tau)$ satisfying Eqs. (13)–(16), and use these functions to express q'_i,p'_i as functions of σ,τ. Consider the action S corresponding to the region D between the lines $\tau = \tau_1$ and $\tau = \tau_2$ [see (18)]. Note that

$$\begin{aligned} S - \iint_D d\sigma \wedge d\tau\, p'_i \frac{\partial q'^i}{\partial \tau} &= \iint_D d\sigma \wedge d\tau \frac{\delta T}{\delta q^i(\sigma)}\frac{\partial q^i}{\partial \tau} + \int_{\tau_1}^{\tau_2} d\tau \frac{\partial T}{\partial \tau} \\ &\quad + \iint_D d\sigma \wedge d\tau \frac{\delta T}{\delta q'^i(\sigma)}\frac{\partial q'^i}{\partial \tau} \\ &= [T]_{\tau_1}^{\tau_2}, \end{aligned} \tag{29}$$

so that a variation of the functions $q^i(\sigma,\tau)$, $p_i(\sigma,\tau)$ will produce a change

$$\begin{aligned} \delta S - \delta \iint_D d\sigma \wedge d\tau\, p'_i \frac{\partial q'^i}{\partial \tau} &= [\delta T]_{\tau_1}^{\tau_2} \\ &= \left[\int d\sigma \left(p_i - \pi_{ij}\frac{\partial q^j}{\partial \sigma}\right)\delta q^i\right]_{\tau_1}^{\tau_2} - \left[\int d\sigma\, p'_i \delta q'^i\right]_{\tau_1}^{\tau_2}. \end{aligned} \tag{30}$$

But we know from (19) that the first term on the right-hand side is equal to δS, contributions from the extremities of the string being ignored once again.[5] So the remaining terms must be equal, i.e.,

$$\iint_D d\sigma \wedge d\tau \left(\delta p'_i \frac{\partial q'^i}{\partial \tau} + p'_i \frac{\partial \delta q'^i}{\partial \tau}\right) = \iint_D d\sigma \wedge d\tau \frac{\partial}{\partial \tau}(p'_i \delta q'^i), \tag{31}$$

i.e.,

$$\iint_D d\sigma \wedge d\tau \left(\delta p'_i \frac{\partial q'^i}{\partial \tau} - \delta q'^i \frac{\partial p'_i}{\partial \tau}\right) = 0. \tag{32}$$

Since p'_i and q'^i can be varied independently of each other (though not of q, p), we conclude that

$$\frac{\partial q'^i}{\partial \tau} = 0, \quad \frac{\partial p'_i}{\partial \tau} = 0. \tag{33}$$

Thus we have obtained constant τ-dependent functionals of q, p. If q, p are expressed in terms of q', p', and τ, the general motion of the string is obtained. Our HJ equation can thus be used exactly as in ordinary mechanics. Note that in the transformed coordinates, we have only q' and p': there is no ϕ or π to complicate the situation.

Note added in proof: We wish to point out that $p_i - \pi_{ij}\frac{\partial q^j}{\partial \sigma}$ is the ith canonical momentum and $H - \phi_i \frac{\partial q^i}{\partial \sigma}$ the Hamiltonian density, so that our HJ system is nothing but the field-theory version of (25).

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).

[2] Y. Nambu, Phys. Lett. B **92**, 327 (1980).

[3] This peculiarity, of course, is not to be regarded as a defect of Nambu's mechanics. In fact it is quite a standard feature of string dynamics: cf. C. Rebbi, Phys. Rep. C **12**, 1 (1974). The treatment of time as something "external"—the way it is done in Newtonian mechanics—is not very suitable for relativistic theories. However, our interest is not so much in string systems as in possible generalizations of Hamiltonian mechanics.

[4] In order to make the summation convention apply to terms containing π_{ij} and related quantities, we formally define operations like $\partial/\partial\pi_{ij}$ for all values i and j by antisymmetry. Strictly speaking, only $n(n-1)/2$ of these make sense, for there are only so many independent π_{ij}.

[5] The extremities of the string do not contribute if they are fixed, so that $\delta q^i = 0$. If the string is closed, then again there is no contribution because the two extremities cancel each other (one might also say that there is no extremity). In the case of an infinite string, too, the "ends" may presumably be ignored.

[6] Note that this need not be the most general solution.

Sine-Gordon and modified Korteweg–de Vries charges

K. M. Case and A. M. Roos
Center for the Study of Nonlinear Dynamics, La Jolla Institute, P.O. Box 1434, La Jolla, California 92038 and The Rockefeller University, New York, New York 10021

(Received 6 October 1981; accepted for publication 25 November 1981)

The sine-Gordon and the integrated modified Korteweg–de Vries equations are shown to conserve the same infinite set of charges. The charges are determined by a recursion relation. As a consequence, the solutions of all the equations generated by the charges have in common all time-independent properties.

PACS numbers: 03.40. – t, 03.65.Fd

I. INTRODUCTION

In this paper it is shown that the sine-Gordon and one form of integrated modified Korteweg–de Vries (IMKdV) equations conserve the same set of charges.[1] The set is generated by a recursion relation involving powers of an operator $\mathscr{L}$. With the relation, it is easily verified that the set is in involution. The charges will also be referred to as Hamiltonians, since each one determines an evolution equation. The IMKdV Hamiltonian and the higher order Hamiltonians usually associated with that equation come from the positive powers of $\mathscr{L}$. The inverse powers of $\mathscr{L}$ yield the sine-Gordon Hamiltonian and the higher order Hamiltonians described as the dual sequence in Ref. 2. By using the recursion relation, Lax[3] pairs are easily constructed for the entire set of Hamiltonian equations. The identical linear scattering problem is associated with each equation. Thus, the solutions of these equations have in common all time-invariant properties.

The recursion relation is derived in Sec. II. Several low-order charges are explicitly constructed, and an algorithm is given for evaluating the inverse powers of $\mathscr{L}$, in Sec. III. The Lax pairs are derived in Sec. IV. Throughout the paper the analogous results for sinh-Gordon and the other form of the IMKdV equations are indicated.

II. RECURSION RELATION

The sine-Gordon equation in light cone coordinates, $u_{xt} = \sin u$, the sinh–Gordon equation, $u_{xt} = \sinh u$, and the two forms of the IMKdV equation $u_t = +u_{xxx}(\pm)u_x^3/2$, can be written in Hamiltonian form

$$u_t = \{u, H\}, \tag{1}$$

where the Lie bracket is

$$\{F, G\} = \int_{-\infty}^{+\infty} dx \, \frac{\delta F}{\delta u(x)} L \frac{\delta G}{\delta u(x)}. \tag{2}$$

The symplectic operator is defined by

$$L\phi \equiv \partial_x^{-1}\phi = \frac{1}{2}\left(\int_{-\infty}^{x} \phi \, dy - \int_{x}^{+\infty} \phi \, dy\right). \tag{3}$$

Thus, with the bilinear form

$$\langle F, G\rangle \equiv \int_{-\infty}^{\infty} FG \, dx, \tag{4}$$

the bracket can be written

$$\{F, G\} = \left\langle \frac{\delta F}{\delta u}, \frac{L\delta G}{\delta u} \right\rangle. \tag{5}$$

The appropriate Hamiltonians are

$$\begin{aligned} &\text{sine-Gordon:} && \int_{-\infty}^{\infty} (1 - \cos u)\, dx, \\ &\text{sinh-Gordon:} && \int_{-\infty}^{\infty} (1 - \cosh u)\, dx, \\ &(\pm)\text{IMKdV:} && \int_{-\infty}^{\infty} (u_{xx}^2/2(\mp)u_x^4/8)\, dx. \end{aligned} \tag{6}$$

Notice that a pure imaginary solution of the sine-Gordon equation gives a pure real solution of the sinh-Gordon equation and vice versa. Also, real (imaginary) solutions of + IMKdV yield imaginary (real) solutions of – IMKdV. It is the sine-Gordon and + IMKdV equations and the sinh-Gordon and – IMKdV equations that conserve the same set of charges. The results will be derived for the sine-Gordon, + IMKdV set. Since the results do not depend on reality conditions they are also valid for the sinh-Gordon, – IMKdV set of charges.

It is clear from the form of bracket (2) that the quantity $P = \int_{-\infty}^{\infty}(-u_x^2/2)\, dx$ generates translations and is the conserved momentum for these dynamical systems. From dual Hamiltonian work,[4] one suspects that it is possible to construct a symplectic operator M such that the above dynamical equations are given by another Lie bracket and different but related Hamiltonians.

For example, the momentum P is the + IMKdV Hamiltonian if

$$\text{(a)}\ M\frac{\delta P}{\delta u} = Mu_{xx} = +u_{xxx} + \frac{u_x{}^3}{2} = L\frac{\delta H_{\text{IMKdV}}}{\delta u}, \tag{7}$$

$$\text{(b) the bracket } \{F, G\}_M \equiv \langle F, MB\rangle \tag{8}$$

is antisymmetric and satisfies the Jacobi identity. These conditions are met by

$$M\phi \equiv \partial_x\phi + u_x\partial_x{}^{-1}(u_x\phi). \tag{9}$$

With this bracket, the sine-Gordon Hamiltonian generates translations. Since

$$M\frac{\delta H_{s-G}}{\delta u} = M(\sin u) = u_x = L\frac{\delta P}{\delta u}. \tag{10}$$

It will be clear which charge then generates the sine-Gordon evolution, once the recursion relation is established.

 0022-2488/82/030392-04$02.50 © 1982 American Institute of Physics

The second symplectic operator for the sinh-Gordon, − IMKdV case is

$$M\phi = \partial x\phi - u_x\partial_x{}^{-1}(u_x\phi). \tag{11}$$

By introducing the following highly suggestive notation

$$H_{s-G} \equiv H^{(-1)}, \quad P \equiv H^{(0)}, \quad H_{-\,\mathrm{IMKdV}} \equiv H^{(1)} \tag{12}$$

and

$$\frac{\delta H^{(\pm n)}}{\delta u} \equiv Q^{(\pm n)}, \tag{13}$$

Eqs. (7) and (10) are

$$\begin{aligned} MQ^{(0)} &= LQ^{(1)}, \\ MQ^{(-1)} &= LQ^{(0)}. \end{aligned} \tag{14}$$

$\mathscr{L} \equiv L^{-1}M$ emerges as the lifting operator for the gradients of the conserved charges. The $H^{(\pm n)}$ determined by

$$\mathscr{L}^{\pm n}Q^{(0)} = Q^{(\pm n)} \tag{15}$$

are conserved by the sine-Gordon and (+)IMKdV equations. In fact, the $H^{(\pm n)}$ are in involution.[5]

This is readily shown since the operators L and M are antisymmetric. Let m, n be positive or negative integers such that $m > n$. Then,

$$\begin{aligned} \{H^{(m)}, H^{(n)}\} &= \langle Q^{(m)}, LQ^{(n)}\rangle \\ &= -\langle LQ^{(m)}, Q^{(n)}\rangle \\ &= -\langle MQ^{(m-1)}, Q^{(n)}\rangle \\ &= \langle Q^{(m-1)}, MQ^{(n)}\rangle \\ &= \langle Q^{(m-1)}, LQ^{(n+1)}\rangle \\ &= \{H^{(m-1)}, H^{(n+1)}\}. \end{aligned} \tag{16}$$

Repeating these steps r times yields

$$\{H^{(m)}, H^{(n)}\} = \{H^{(m-r)}, H^{(n+r)}\}. \tag{17}$$

Assuming $m > n$, there are two alternatives: r can be picked such that $m = n + 2r$ and

$$\{H^{(m)}, H^{(n)}\} = \{H^{(n+r)}, H^{(n+r)}\} = 0 \tag{18}$$

or $m = n + 2r - 1$ and

$$\begin{aligned} \{H^{(m)}, H^{(n)}\} &= \{H^{(n+r-1)}, H^{(n+r)}\} \\ &= \langle Q^{(n+r-1)}, MQ^{(n+r-1)}\rangle \\ &= -\langle MQ^{(n+r-1)}, Q^{(n+r-1)}\rangle \\ &= -\langle LQ^{(n+r)}, Q^{(n+r-1)}\rangle \\ &= \{H^{(n+r)}, H^{(n+r-1)}\} = 0. \end{aligned} \tag{19}$$

For $m < n$, one obtains the same results, using the antisymmetry of the bracket.

Thus for all m, n

$$\{H^{(m)}, H^{(n)}\} = 0. \tag{20}$$

III. CONSTRUCTION OF CHARGES

The construction of the positive $H^{(n)}$'s is straightforward. For example,

$$\begin{aligned} Q^{(1)} &= \mathscr{L}Q^{(0)} = \partial_x^2 Q^{(0)} + \partial_x(u_x\partial_x^{-1}(u_xQ^{(0)})) \\ &= \partial_x\left(u_{xxx} + \frac{u_x{}^3}{2}\right), \end{aligned} \tag{21}$$

$$Q^{(2)} = \mathscr{L}Q^{(1)} = \partial_x(u_{xxxxx} + \tfrac{3}{8}u_x^5 + \tfrac{5}{2}(u_xu_{xx}{}^2 + u_x^2u_{xxx})). \tag{22}$$

Therefore,

$$H^{(1)} = \int_{-\infty}^{\infty}\left(\frac{u_{xx}{}^2}{2} - \frac{u_x^4}{8}\right)dx, \tag{23}$$

$$H^{(2)} = \int_{-\infty}^{\infty}\left(\frac{-u_{xxx}{}^2}{2} - \frac{u_x^6}{16} + \frac{5}{4}u_x^2u_{xx}{}^2\right)dx. \tag{24}$$

The construction of the negative charges is more difficult since it involves inverting $\mathscr{L}$. However, they can be determined by an alternative method that exploits the symmetry in the variables x and t, of the sine-Gordon equation. This elucidates the relation of the positive charges to the negative ones.

For each sine-Gordon conservation law

$$\partial_t j_0 + \partial_x j_1 = 0 \tag{25}$$

there is a corresponding law

$$\partial_t j_1' + \partial_x j_0' = 0, \tag{26}$$

where the prime denotes interchanging the x and t derivatives and then eliminating the t derivatives using the equation of motion. For example, the sine-Gordon energy density is the charge density corresponding to the momentum flux density and vice versa.

$$\partial_t(1 - \cos u) = \sin u\, u_t = u_{xt}u_t = +\partial_x(u_t^2/2). \tag{27}$$

$$\partial_t(-u_x^2/2) = -u_xu_{xt} = -u_x\sin u = \partial_x(\cos u - 1). \tag{28}$$

As a second example, consider the (+)IMKdV energy density

$$j_0 = \frac{u_{xx}{}^2}{2} - \frac{u_x^4}{8}. \tag{29}$$

Then,

$$\partial_t j_0 = +u_{xx}u_{xxt} - \frac{u_x{}^3}{2}u_{xt} \tag{30}$$

$$= \partial_x\left(\frac{u_x{}^2}{2}\cos u\right). \tag{31}$$

Thus, the corresponding charge density is

$$j_1' = -\frac{(\partial_x{}^{-1}\sin u)^2}{2}\cos u. \tag{32}$$

Therefore the charge is

$$H' = \int_{-\infty}^{\infty}dx\left[-\cos u\,\frac{(\partial_x{}^{-1}\sin u)^2}{2}\right] \tag{33}$$

and the gradient of the charge is

$$Q' = +\sin u\,\frac{[\partial_x{}^{-1}\sin u]^2}{2} + \cos u\partial_x{}^{-1}(\cos u\partial_x{}^{-1}\sin u). \tag{34}$$

Through successive integration by parts and careful evaluation of the limits one finds $MQ' = LQ^{(-1)}$. Thus, $Q^{(-2)} = Q'$, $H^{(-2)} = H'$. $H^{(-2)}$ is the sine-Gordon Hamiltonian when M defines the Lie bracket.

Clearly, the set of charges corresponding to those determined by positive powers of $\mathscr{L}$ is not some new set, but those

generated by the inverse powers of $\mathscr{L}$. The symmetry in x and t yields an algorithm for evaluating $\mathscr{L}^{-n}$. On interchanging x and t and using $u_t = \partial x^{-1} \sin u$, the flux densities of the charges obtained from the positive powers of $\mathscr{L}$ become the charge densities determined by the negative powers of $\mathscr{L}$. For example, $H^{(-3)}$ is constructed from $H^{(2)}$ in this way

$$\partial_t\left(\frac{-u_{xxx}^2}{2} - \frac{u_x^6}{16} + \frac{5}{4}u_x^2 u_{xx}^2\right) = \partial_x\left[\left(\frac{3}{8}u_x^4 - \frac{u_{xx}^2}{2}\right)\cos u + u_{xx}u_x^2 \sin u\right], \tag{35}$$

from which it follows that

$$H^{(-3)} = \int_{-\infty}^{\infty} dx\left[-\sin u\{\partial_x^{-1}[\cos u\partial_x^{-1}\sin u](\partial_x^{-1}\sin u)^2\} + \cos u\{3/8(\partial_x^{-1}\sin u)^4 - 1/2(\partial_x^{-1}[\cos u\partial_x^{-1}\sin u])^2\}\right] \tag{36}$$

and

$$\begin{aligned} Q^{(-3)} = & -\sin u\{-3/8(\partial_x^{-1}\sin u)^4 + 1/2(\partial_x^{-1}[\cos u\partial_x^{-1}\sin u])^2 \\ & + \partial_x^{-1}\sin u\{\partial_x^{-1}[\cos u\partial_x^{-1}[\cos u\partial_x^{-1}\sin u]] + \partial_x^{-1}[\sin u(\partial_x^{-1}\sin u)^2]\}\} \\ & + \cos u\{3/2\partial_x^{-1}[\cos u(\partial_x^{-1}\sin u)^3] - \partial_x^{-1}[\cos u\{\partial_x^{-1}[\sin u(\partial_x^{-1}\sin u)^2] \\ & + \partial_x^{-1}[\cos u\partial_x^{-1}[\cos u\partial_x^{-1}\sin u]]\}] - (\partial_x^{-1}\sin u)^2\partial_x^{-1}[\cos u\partial_x^{-1}\sin u] \\ & + 2\partial_x^{-1}[\sin u(\partial_x^{-1}\sin u)\partial_x^{-1}[\cos u\partial_x^{-1}\sin u]]\}. \end{aligned} \tag{37}$$

Again, one can verify that $LQ^{(-2)} = MQ^{(-3)}$. Naturally, these charges can also be constructed using Bäcklund transformations, as shown in the Appendix.

IV. LAX PAIRS

The information needed to construct solutions to a nonlinear evolution equation is contained in two associated linear problems, the so-called Lax pair. With the recursion relation (15) it is easy to construct a Lax pair for each equation,

$$u_t^{(n)} = \{u^{(n)}, H^{(n)}\}, \quad n = \pm 1, \pm 2\ldots. \tag{38}$$

Recall that a Lax pair is an eigenvalue equation and a time-dependent equation whose compatibility condition (assuming the time independence of the eigenvalues) is that the nonlinear equation be satisfied. If the Lax pair is

$$\lambda\psi = \tilde{\mathscr{L}}\psi, \tag{39a}$$

$$\partial_t\psi = B^{(n)}\psi, \tag{39b}$$

then the compatibility condition is

$$\partial_t\tilde{\mathscr{L}} = [B^{(n)}, \tilde{\mathscr{L}}], \tag{40}$$

where the bracket is the commutator.

To construct (39a), introduce the function

$$\psi \equiv \sum_{n=-\infty}^{+\infty} \frac{Q^{(n)}}{\lambda^n}, \tag{41}$$

where λ is a constant. Equation (15) can be written

$$\mathscr{L}Q^{(n)} = Q^{(n+1)}. \tag{42}$$

Multiplying both sides of (42) by λ^{-n} and then summing gives the eigenvalue equation

$$\lambda\psi = \mathscr{L}\psi. \tag{43}$$

The operator $\tilde{\mathscr{L}}$ in (39a) is $\mathscr{L}$ for all the equations (38). Thus, one-half of the Lax pair is constructed.

Since only the time dependent part of the pair differs for each equation, all time independent properties are shared by the solutions.

The time dependent problem (39b) is the linearized evolution equation. Write the evolution equation as

$$u_t = K[u], \tag{44}$$

where K is some functional. Then, linearized K is defined by

$$K'[\Delta u] \equiv \lim_{\epsilon\to 0}\frac{K[u+\Delta u] - K[u]}{\epsilon}. \tag{45}$$

The linear equation,

$$\partial_t\Delta u = K'[\Delta u] \tag{46}$$

is satisfied by

$$\Delta u = \{u, G\}, \tag{47}$$

where G is some conserved functional.[6] For

$$\Delta u = \{u, H^{(n)}\} = LQ^{(n)}, \tag{48}$$

Eq. (46) becomes

$$\partial_t LQ^{(n)} = K'[LQ^{(n)}]. \tag{49}$$

Multiplying both sides of this by λ^{-n} and summing gives

$$\partial_t\psi = L^{-1}K'[L\psi]. \tag{50}$$

Thus, $B^{(n)}$ can be constructed for every evolution equation.

As an example, the linear operator for the first sine-Gordon Hamiltonian is

$$B^{(-1)}\psi = \cos u\partial_x^{-1}\psi. \tag{51}$$

One can easily verify that

$$\partial_t = [B^{(-1)}, \mathscr{L}] \tag{52}$$

implies $u_{xt} = \sin u$.

Note that the pair constructed here is not in the standard form, which is

$$\zeta\begin{pmatrix} v_1 \\ v_x \end{pmatrix} = i\begin{pmatrix} \partial_x & -q \\ r & -\partial_x \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathscr{L}^s\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \tag{53a}$$

$$\frac{\partial}{\partial t}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & -A \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = B^s\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \tag{53b}$$

with

$$r = -q = u_x/2,\ A = \frac{i\cos u}{4\zeta},\ B = C = \frac{i\sin u}{4\zeta}. \quad (54)$$

The operators $(\mathscr{L}, B^{(-1)})$ are equivalent to the standard ones $(\mathscr{L}^s, B^s)$ for

$$\lambda = -4\zeta^2 \text{ and } \psi = (v_1^2 - v_2^2). \quad (55)$$

Analogous results hold for the sinh-Gordon, $(-)$IMKdV hierarchy of equations with the appropriate definition (11) of $\mathscr{L}$. For sinh-Gordon, $B^{(-1)}\psi = \cosh u\partial_x^{-1}\psi$. The standard form of the Lax pair is given by (44a) and (44b) with

$$r = q = \frac{u_x}{2},\ A = \frac{i\cosh u}{4\zeta},\ -B = C = \frac{i\sinh u}{4\zeta}. \quad (56)$$

This is equivalent to $(\mathscr{L}, B^{(-1)})$ if

$$\lambda = -4\zeta^2,\quad \psi = (v_1^2 + v_2^2). \quad (57)$$

V. CONCLUSIONS

It is remarkable that equations of the higher order sine (sinh)-Gordon Hamiltonians form part of the hierarchy $+(-)$IMKdV equations. Because there is a recursion relation that determines these charges, the identical linear eigenvalue problem is associated with all the equations generated by them. The consequences are that for these equations

1. The time-independent half of the Bäcklund transformations is the same.
2. The Marchenko equations are the same.
3. The nonlinear superposition laws are the same.
4. Given a solution to one of the equations, for example, an n solition, one obtains a corresponding solution to any of the other equations by appropriately modifying the time dependence.

ACKNOWLEDGMENTS

This work was supported in part by the Department of Energy under Contract No. DE-AC02-81ER 10807 and in part by the Office of Naval Research under Contract No. N-00014-79-C-0537.

A. M. Roos wishes to thank Björn Birnir for conversations which were a great help and fun.

APPENDIX

The conserved densities can also be constructed from the sine-Gordon Bäcklund transformation

$$1/2(u - v)_x = \gamma \sin 1/2(u + v), \quad (A1)$$

$$1/2(u + v)_t = (\gamma)^{-1}\sin 1/2(u - v). \quad (A2)$$

One uses the symmetry in the variables. This expression, on interchanging x and t, is again a sine-Gordon Bäcklund transformation. The charge densities calculated from (A1) and (A2) are those generated by the negative powers of $\mathscr{L}$; on interchanging the variables, the transformation yields the charges generated by the positive powers.

The charge and current densities are determined as follows: multiplying the first half of the transformation by $1/2(u + v)_t$ and the second half by $1/2(u - v)_x$ and subtracting yields

$$\gamma^2(u + v)_t \sin 1/2(u + v) - (u - v)_x \sin 1/2(u - v) = 0, \quad (A3)$$

which implies

$$-\gamma^2[\cos 1/2(u + v)]_t + [\cos 1/2(u - v)]_x = 0. \quad (A4)$$

Consider γ as a small parameter and expand u near $\gamma = 0$,

$$u = u_0 + \gamma u_1 + \gamma^2 u_2 + \cdots. \quad (A5)$$

Substituting this into (A4) gives a conservation law for each order in γ. As illustration, to second order in γ, (A4) is

$$(1 - \cos v)_t + 1/2(v_t^2)_x = 0. \quad (A6)$$

The odd powers in γ yield terms which are merely total time derivatives of the previous terms.

The u_i for $i > 0$ are found as functions of u_0 from the expansion of (A2). $u_0 = v$ comes from the 0th order of (A4). For example,

$$\begin{aligned} u_1 &= 2u_{0t} = 2v_t, \\ u_2 &= 2v_{tt}, \\ u_3 &= 2v_{ttt} + v_t^3/3, \text{ etc.} \end{aligned} \quad (A7)$$

The higher order charge densities derived in this way differ from those given in the text by terms which are total time derivatives.

[1] The various original derivations of the modified KdV and the sine-Gordon charges are in: A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, Proceedings of the IEEE **61**, 1443 (1973); M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974); M. D. Krusdal and D. Wiley, Am. Math. Soc. *Summer Seminar on Nonlinear Wave Motion*, edited by A. C. Newell (Am. Math. Soc., Providence, R. I., 1972).

[2] B. Yoon, Phys. Rev. D **13**, 3440 (1976).

[3] P. D. Lax, Comm. Pure Appl. Math. **21**, 647–690 (1968).

[4] F. Magri, J. Math. Phys. **19**, 1156 (1978).

[5] With either bracket (2) or (8) the $H^{(\pm n)}$, $n = 0, 1, 2, \ldots$ form an abelian Lie algebra.

[6] K. M. Case, Phys. Rev. Lett. **40**, 351 (1978).

Inverse problems for nonabsorbing media with discontinuous material properties

Robert J. Krueger
Applied Mathematical Sciences, Ames Laboratory-USDOE, Iowa State University, Ames, Iowa 50011 [a)]

(Received 8 May 1981; accepted for publication 11 September 1981)

One-dimensional electromagnetic and elastic inverse problems are formulated for media with discontinuous material properties. In addition, an impedence mismatch between source and medium is allowed. The measured data for either problem is shown to generate a reflection kernel which is used in the solution of the inverse problem. The solution algorithm itself is a time domain technique which is a special case of previously obtained results for absorbing media.

PACS numbers: 03.40.Dz, 03.50.De

1. INTRODUCTION

The problem of reconstructing the material properties of a medium from data collected outside of the medium is of central importance in disciplines ranging from medicine and engineering sciences to the geosciences. For example, scattered electromagnetic waves can be used to reconstruct the permittivity and conductivity profiles of a one-dimensional medium, while stress impulse response data can determine the characteristic impedence of an elastic medium. A shortcoming of many of the techniques developed for such inverse problems is that it is assumed the properties of the medium do not suffer jump discontinuities. Such an assumption is frequently ill founded and, indeed, can complicate the data required in a reconstruction algorithm as a comparison of the result of Krueger[1,2] and Weston[3] shows.

In previous work,[1,2] hereafter referred to as I, II, an inverse problem was studied for a one-dimensional absorbing medium with discontinuous material properties. The motivation for that work was the electromagnetic inverse problem mentioned above. The scatterer was assumed to be of finite depth so that transmitted as well as reflected waves could be used in the inversion process. A time domain approach to the problem generated a Gel'fand–Levitan type of integral equation whose kernel and forcing term are constructed from the transmission and reflection data. The solution of that integral equation is used to generate the permittivity and conductivity profiles of the scattering medium. It was also shown how the results specialize for the case of piecewise constant profiles.

The purpose of this paper is to:

1. show how the results of I and II simplify in the nonabsorptive case;
2. show how the nonabsorptive version of I and II can be used in an inverse problem for elastic media, namely, determining the characteristic impedence of the medium.

Three important properties of these results are:

1. Discontinuities in the material properties of the medium are allowed. This also implies that there can be an impedence mismatch between source and medium.
2. For the nonabsorptive case, reflection data alone suffices to reconstruct the properties of the scatterer. Hence, a semi-infinite model of the medium may be used.
3. For the case in which there are no discontinuities, the results of this paper reduce to the Marchenko approach.

In Sec. 2 the formulation of the electromagnetic inverse problem is reviewed and specialized to the case of a nonabsorbing medium. Section 3 formulates the inverse problem for a one-dimensional elastic medium. Section 4 provides a solution technique for these inverse problems for a single layer medium. The data required for this solution technique is discussed in Sec. 5. Section 6 presents a solution algorithm for a multilayered medium, and examples are given in Sec. 7. These results are summarized and discussed in Sec. 8.

2. FORMULATION OF THE ELECTROMAGNETIC INVERSE PROBLEM

The equation modeling one-dimensional electromagnetic wave propagation in a nonmagnetic, nonabsorbing medium is

$$E_{zz} - \mu_0 \epsilon(z) E_{tt} = 0, \quad -\infty < z < \infty, \tag{2.1}$$

where $E(z,t)$ is the electric field, $\epsilon(z)$ is the permittivity, and μ_0 is the (constant) permeability. Assume that $\epsilon(z) = \epsilon_0$ for $z < 0$ and that for $z > 0$, $\epsilon(z)$ and its first two derivatives are piecewise continuous; i.e., $\epsilon(z)$ is a C^2 function for $z > 0$ except for a finite number of jump discontinuities at $z_0, z_1, z_2, \ldots$, where $0 = z_0 < z_1 < z_2 < \cdots$. In particular, a medium of finite depth can be considered by making $\epsilon(z)$ constant for $z > z_n$ for some n.

Assume that $E(z,t) \equiv 0$ for $z > 0$, $t < 0$ and that an incident field $E^i(z - ct)$ (where $c = 1/(\mu_0 \epsilon_0)^{1/2}$) propagating along the negative z axis is incident normally on the medium at time $t = 0$. Thus $E^i(z - ct) \equiv 0$ if $z - ct > 0$ and the initial data for Eq. (2.1) is

$$E(z,t) = E^i(z - ct), \quad -\infty < z < \infty, \ t < 0. \tag{2.2}$$

Let E^r denote the corresponding reflected wave; i.e.,

$$E(z,t) = E^r(z + ct), \quad z < 0, \ t > 0,$$

where $E^r(z + ct) = 0$ if $z + ct < 0$. Finally, continuity of the

[a)] Permanent address: Department of Mathematics and Statistics, University of Nebraska, Lincoln, Nebraska 68588.

 0022-2488/82/030396-09$02.50 © 1982 American Institute of Physics

electric and magnetic fields implies that

$$E(z_i - ,t) = E(z_i + ,t) \tag{2.3a}$$

and

$$E_z(z_i - ,t) = E_z(z_i + ,t) \tag{2.3b}$$

for $i = 0,1,2,\cdots$.

The inverse problem for Eq. (2.1) can be stated as follows: Given the incident wave $E^i(t)$ for $t<0$ and the corresponding reflected wave $E^r(t)$ for $t>0$, determine the permittivity $\epsilon(z)$ of the medium. (Figure 1 illustrates the situation.) Before proceeding to the solution of this inverse problem, Eq. (2.1) will be rewritten in terms of a travel time coordinate x, where

$$x = x(z) = \int_0^z [\mu_0 \epsilon(s)]^{1/2} ds, \tag{2.4}$$

$$x_i = x(z_i), \quad i = 0,1,2,\cdots$$

and

$$u(x,t) = E(z,t).$$

Then the initial value problem (2.1)–(2.3) takes the form

$$u_{xx} - u_{tt} + A(x)u_x = 0, \quad -\infty < x < \infty, \tag{2.5}$$

$$u(x,t) = u^i(x-t), \quad -\infty < x < \infty, \ t<0, \tag{2.6}$$

$$u(x_i - ,t) = u(x_i + ,t), \tag{2.7a}$$

$$[\epsilon(z_i -)]^{1/2} u_x(x_i - ,t) = [\epsilon(z_i +)]^{1/2} u_x(x_i + ,t), \tag{2.7b}$$

where

$$A(x) = -\frac{d}{dz}[\epsilon(z)\mu_0]^{1/2}, \quad x \neq x_i \tag{2.8}$$

and

$$u^i(t) = E^i(ct). \tag{2.9}$$

In the transformed problem, the reflected wave u^r is given by

$$u^r(t) = E^r(ct). \tag{2.10}$$

Notice that the curved characteristics of Eq. (2.1) become straight lines for Eq. (2.5) because of the transformation (2.4). This is illustrated in Fig. 2.

The inverse problem for Eq. (2.5) is to determine $A(x)$ from the data $u^i(t)$, $u^r(t)$. In particular, it will be shown in Secs. 4 and 6 that if $u^i(-t)$ and $u^r(t)$ are given for $0<t<t_m$ for some t_m, then $A(x)$ can be determined for $0<x<t_m/2$ and $\epsilon(z)$ can be determined in some finite interval $0<z<z_m$.

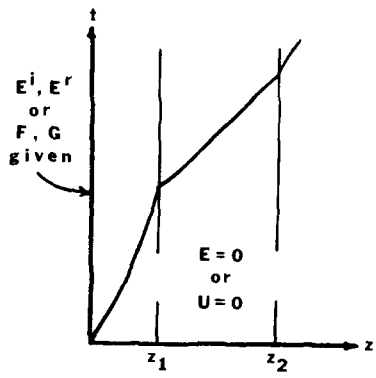

FIG. 1. The given data are used to reconstruct ϵ [in Eq. (2.1)] or ρc [from Eq. (3.1)] in the region $z>0$. The characteristic curves of (2.1) and (3.1) separate the quiescent and active regions.

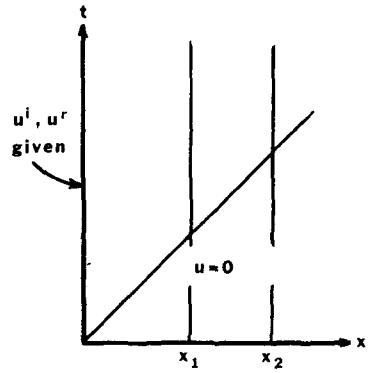

FIG. 2. The given data are used to reconstruct $A(x)$ [in Eq. (2.5)] in the region $x>0$. The characteristic line $x=t$ separates the quiescent and active regions for problem (2.5)–(2.10).

3. FORMULATION OF THE ELASTIC INVERSE PROBLEM

The equation modeling one-dimensional wave propagation in an isotropic inhomogeneous elastic medium is

$$(\rho c^2 U_z)_z = \rho U_{tt}, \quad z>0, \ t>0, \tag{3.1}$$

where $U(z,t)$ is the elastic displacement, $\rho(z)$ is the density of the medium, and $c(z)$ is the wave speed. Thus

$$\rho c^2 = \mu$$

or

$$\rho c^2 = \lambda + 2\mu,$$

depending on whether U denotes transverse or longitudinal displacement, respectively. The Lamé parameters μ and λ are considered to be functions of z. The functions ρ and c are assumed to be C^2 for $z>0$ except for jump discontinuities at $z_1, z_2, z_3, \ldots$, where $0<z_1<z_2<\cdots$.

Assume that initially there is no disturbance in the medium and that beginning at $t=0$, a prescribed stress is imposed along the boundary $z=0$. The source of this stress may be modeled by an elastic wave equation in the region $z<0$, $(\rho c^2 U_z)_z = \rho U_{tt}$, where now ρ and c are assumed constant and known. Thus, the initial-boundary conditions for Eq. (3.1) are

$$U(z,0) = U_t(z,0) = 0, \quad z>0 \tag{3.2}$$

and

$$(\rho c^2)|_{z=0-} U_z(0-,t) = F(t), \quad t>0, \tag{3.3}$$

where $F(t)$ is the prescribed stress and $(\rho c^2)|_{z=0-}$ is the modulus of elasticity of the source. Notice that the boundary condition (3.3) allows for an impedence mismatch along $z=0$. Also, continuity of displacement and stress is imposed at the interfaces $0=z_0, z_1, z_2, \ldots$ so that

$$U(z_i - ,t) = U(z_i + ,t) \tag{3.4a}$$

and

$$(\rho c^2)|_{z=z_i-} U_z(z_i - ,t) = (\rho c^2)|_{z=z_i+} U_z(z_i + ,t) \tag{3.4b}$$

for $i = 0, 1, 2, \cdots$. Finally, let $G(t)$ denote the resulting velocity at $z = 0$,

$$U_t(0,t) = G(t). \tag{3.5}$$

Thus, for example, if $F(t) = \delta(t)$, where δ is the Dirac delta, then $G(t)$ is the impulse response velocity field of the medium. [Notice that continuity of U across $z = 0$ implies that a one-sided limit is not needed in (3.5).]

The inverse problem for Eq. (3.1) can be stated as follows: Given the functions $F(t)$ and $G(t)$ for $t > 0$, determine the impedence ρc of the medium for $z > 0$. (Figure 1 again illustrates the situation.) In order to cast this problem in a more familiar setting, a travel time coordinate x is introduced:

$$x = x(z) = \int_0^z [c(s)]^{-1}\, ds, \tag{3.6}$$

$$x_i = x(z_i), \quad i = 0,1,2,\cdots$$

and

$$u(x,t) = U(z,t).$$

Thus, the initial-boundary value problem (3.1)–(3.4) takes the form

$$u_{xx} - u_{tt} + A(x)u_x = 0, \quad x > 0, \tag{3.7}$$

$$u(x,0) = u_t(x,0) = 0, \quad x > 0, \tag{3.8}$$

$$(\rho c)|_{x=0-}\, u_x(0-,t) = F(t), \quad t > 0, \tag{3.9}$$

$$u(x_i - ,t) = u(x_i + ,t), \tag{3.10a}$$

$$(\rho c)|_{z=z_i-}\, u_x(x_i - ,t) = (\rho c)|_{z=z_i+}\, u_x(x_i + ,t), \tag{3.10b}$$

where

$$A(x) = \frac{d}{dx}\ln(\rho c) = c\frac{d}{dz}\ln(\rho c), \quad x \neq x_i. \tag{3.11}$$

In the transformed problem, the velocity data is unchanged,

$$u_t(0-,t) = G(t). \tag{3.12}$$

The inverse problem for Eq. (3.7) is to determine $A(x)$ from the data $F(t)$, $G(t)$. It will be shown in Secs. 4 and 6 that if $F(t)$ and $G(t)$ are given for $0 < t < t_m$ for some t_m, then $A(x)$ can be determined for $0 < x < t_m/2$ and ρc can be determined as a function of travel time for $0 < x < t_m/2$.

4. SOLUTION OF THE INVERSE PROBLEM IN A SINGLE LAYER

Although the differential equations (2.5) and (3.7) are similar, the nature of the measured data, E^i and E^r vs F and G, is different. For the moment, this difference will be suppressed by assuming that a reflection kernel $R(t)$ can be constructed from either set of data. In this section it is shown how the function $R(t)$ can be used to reconstruct the coefficient $A(x)$ of (2.5) or (3.7) for $0 < x < x_1$. [Section 5 addresses the problem of constructing $R(t)$.]

To streamline the notation in what follows, rewrite the jump conditions (2.7b) and (3.10b) as

$$c_i^-\, u_x(x_i - ,t) = c_i^+\, u_x(x_i + ,t)$$

for $i = 0, 1, 2, \ldots$, where

$$c_i^\pm = [\epsilon(z_1 \pm)]^{1/2}$$

in the electromagnetic case and

$$c_i^\pm = \rho c|_{z=z_i\pm}$$

in the elastic case. In this new notation the constant c_0 in I is given by

$$c_0 = c_0^-/c_0^+. \tag{4.1}$$

To specialize the results of I to nonabsorptive media, note that the functions $B(x)$ and $C(x)$ in I are identically zero for the problems considered in Secs. 2 and 3 of this paper. This implies that the function $G(x)$ in I satisfies $G(x) \equiv 1$. Also, the functions $N_\pm(x,y,0)$ can be shown to satisfy the some characteristic initial value problem,

$$\left[\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} + D(x)\right]N_\pm = 0, \quad 0 < x < x_1, \quad -x < y < x, \tag{4.2}$$

$$N_\pm(x,x,0) = \tfrac{1}{2}\left[(c_0 + 1)\int_0^x D(s)\, ds - A(0)\right], \tag{4.3a}$$

$$N_\pm(x,-x,0) = \tfrac{1}{2}\left[-(c_0 - 1)\int_0^x D(s)\, ds - A(0)\right], \tag{4.3b}$$

where $A(0) = A(0+)$ and

$$D(x) = -\tfrac{1}{4}[(2A'(x) + A^2(x)].$$

It follows by uniqueness that $N_+ = N_-$, so set

$$N(x,y) = N_\pm(x,y,0).$$

Also, identify $R_+(t)$ in I with the function $R(t)$ discussed above.

With these simplifications, Lemma 5 of I yields an integral equation for N,

$$N(x,y) = (c_0 + 1)R(x + y) - f_0 N(x,-y) - \int_{-y}^x R(y + s)N(x,s)\, ds, \tag{4.4}$$

where $f_0 = (c_0 - 1)/(c_0 + 1)$ and $0 < x < x_1$, $-x < y < x$. The solution N of (4.4) is related to $A(x)$ through Eq. (4.3a),

$$2N(x,x) = -A(0) - \tfrac{1}{4}(c_0 + 1)[2\{A(x) - A(0)\} + \int_0^x A^2(s)\, ds] \tag{4.5}$$

or

$$8\frac{d}{dx}N(x,x) = -(c_0 + 1)[2A'(x) + A^2(x)]. \tag{4.6}$$

This, $A(x)$ can be found by solving the Ricatti equation (4.6) for A with initial data

$$A(0) = -2N(0,0).$$

As is discussed elsewhere,[4] the nonlinear Volterra equation (4.5) generally yields more accurate numerical solutions than (4.6).

As it stands, Eq. (4.4) is amenable to numerical solution. However, to cast the equation in a more familiar form, replace y with $-y$ in (4.4) and substitute the resulting expression for $N(x,-y)$ into (4.4) to obtain

$$(c_0 + 1)(1 - f_0^2)N(x,y) = (c_0 + 1)S(x,y) - \int_{-|y|}^x N(x,s)S(s,y)\, ds, \tag{4.7}$$

where $0<x<x_1$, $-x<y<x$ and $S(x,y)=(c_0+1)R(x+y)-(c_0-1)R(x-y)$, with $R(s)=0$ if $s<0$. Equation (4.7) can also be written

$$(c_0+1)(1-f_0^2)N(x,y)=(c_0+1)S(x,y)-\int_{-x}^{x}N(x,y)S(s,y)\,ds. \quad (4.8)$$

Notice that if $c_0=1$, as in the case studied by Gopinath and Sondhi[5,6] and Burridge,[7] then Eq. (4.4) reduces to

$$N(x,y)=2R(x+y)-\int_{-y}^{x}R(y+s)N(x,s)\,ds. \quad (4.9)$$

With the substitution $K(x,y)=-\frac{1}{2}N(x,y)$, Eq. (4.9) is seen to be the usual Marchenko equation.[8]

Once $A(x)$ has been determined for $0<x<x_1$, the permittivity $\epsilon(z)$ in Eq. (2.1) can be determined via the formulas

$$z=\left[c_0\int_0^x\exp\left\{-\int_0^t A(s)ds\right\}dt\right]/[\epsilon_0\mu_0]^{1/2}, \quad (4.10)$$

$$c_0[\epsilon(z)/\epsilon_0]^{1/2}=\exp\left\{\int_0^{x(z)}A(s)\,ds\right\}. \quad (4.11)$$

For the elastic case, the impedence ρc associated with Eq. (3.1) can be determined as a funciton of travel time,

$$\rho c|_x=c_0^+\exp\left\{\int_0^x A(s)\,ds\right\}. \quad (4.12)$$

If $\rho(z)$ is a known constant ρ_0, then c can be determined as a function of z via the formulas

$$z=\left[c_0^+\int_0^x\exp\left\{\int_0^t A(s)\,ds\right\}dt\right]/\rho_0,$$

$$\rho_0 c(z)=c_0^+\exp\left\{\int_0^{x(z)}A(s)\,ds\right\}.$$

Notice also that from Eqs. (3.11) and (4.6) it follows that

$$2\frac{d}{dx}N(x,x)=-(c_0+1)(\rho c|_x)^{-1/2}\frac{d^2}{dx^2}(\rho c|_x)^{1/2}, \quad (4.13)$$

with

$$\rho c|_{x=0+}=c_0\,\rho c|_{x=0-} \quad (4.14)$$

and

$$\frac{d}{dx}\rho c|_{x=0+}=-2N(0,0)\rho c|_{x=0+}. \quad (4.15)$$

When $c_0=1$, this is again the usual Marchenko result for elastic media.

5. CONSTRUCTION OF *R(t)*

The function $R(t)$ used in Eq. (4.4) is the kernel of the reflection operator introduced in I, Eq. (4.12). This implies that the data $u^i(t)$, $u^r(t)$ for the electromagnetic problem is related to $R(t)$ through the equation

$$u^r(t)=f_0u^i(-t)+\int_0^t R(s)u^i(s-t)\,ds,\quad 0<t<2x_1. \quad (5.1)$$

Thus, $R(t)$ is the solution of a Volterra equation of the first kind. If u^i does not contain a delta function singularity, then (5.1) can be solved for $R(t)$ by converting to a Volterra equation of the second kind as shown in I, Sec. 2. The function $R(t)$ can then be constructed uniquely.[9]

As an alternative construction method, Eq. (5.1) can be easily shown to be valid in the case when u^i contains jump discontinuities or delta function singularities. Thus, for example, if

$$u^i(t)=\delta(t), \quad (5.2)$$

then

$$u^r(t)=f_0\delta(t)+R(t). \quad (5.3)$$

Since $R(t)$ is a C^2 function for $0\leqslant t\leqslant 2x_1$, Eq. (5.3) yields both $R(t)$ and f_0 (and therefore c_0).

In order to construct $R(t)$ for the elastic problem, the data $F(t)$, $G(t)$ must be rewritten in terms of right and left moving waves so that Eq. (5.1) can be used. For the sake of comparison, consider how data at $x=0$ in the electromagnetic case can be resolved into right and left moving waves. For $x<0$, the solution of Eq. (2.5) is

$$u(x,t)=u^i(x-t)+u^r(x+t)$$

and so at $x=0-$,

$$u_x(0-,t)=u^{i\prime}(-t)+u^{r\prime}(t),$$

$$u_t(0,t)=-u^{i\prime}(-t)+u^{r\prime}(t).$$

Therefore,

$$u^i(-t)=-\frac{1}{2}\int_0^t[u_x(0-,s)-u_t(0,s)]\,ds$$

and

$$u^r(t)=\frac{1}{2}\int_0^t[u_x(0-,s)+u_t(0,s)]\,ds.$$

Now for the elastic case it follows from (3.9) and (3.12) that

$$u^i(-t)=-\frac{1}{2}\int_0^t\left[\frac{1}{c_0^-}F(s)-G(s)\right]ds, \quad (5.4a)$$

$$u^r(t)=\frac{1}{2}\int_0^t\left[\frac{1}{c_0^-}F(s)+G(s)\right]ds. \quad (5.4b)$$

Substituting Eq. (5.4) into (5.1) again gives a Volterra equation of the first kind for $R(t)$,

$$\int_0^t\left[\frac{1}{c_0^-}(1+f_0)F(s)+(1-f_0)G(s)\right]ds$$

$$=-\int_0^t R(s)\left(\int_0^{t-s}\left[\frac{1}{c_0^-}F(y)-G(y)\right]dy\right)ds,\quad 0<t<2x_1. \quad (5.5)$$

If impulse response data is used, then

$$F(t)=\delta(t). \quad (5.6)$$

In this case, letting $t\to 0$ in Eq. (5.5) yields

$$\lim_{t\to 0+}\int_0^t G(s)\,ds=-\frac{1}{c_0^+}$$

so

$$G(t)=-\frac{1}{c_0^+}\delta(t)+G_c(t), \quad (5.7)$$

where $G_c(t)$ is C^2 for $0\leqslant t\leqslant 2x_1$. Substituting (5.6) and (5.7)

into Eq. (5.5) and differentiating gives a Volterra equation of the second kind,

$$(1-f_0)G_c(t) = -\left(\frac{c_0^+ + c_0^-}{c_0^+ c_0^-}\right)R(t) + \int_0^t R(s) \times G_c(t-s)\,ds, \quad 0<t<2x_1 \tag{5.8}$$

from which $R(t)$ can be constructed uniquely[9] since G_c is C^2.

6. SOLUTION OF THE INVERSE PROBLEM IN A MULTILAYERED MEDIUM

If the coefficient $A(x)$ of Eqs. (2.5) or (3.7) is discontinuous at $x = x_1, x_2, \cdots$ this will, in the case of noiseless data, be apparent from the reflected wave $u^r(t)$ or the response function $G(t)$. For example, if the incident wave is given by

$$u^i(t) = \delta(t),$$

then the reflected wave $u^r(t)$ will contain δ-function singularities at $t = 2x_1$, $t = 2x_2, \ldots$. Of course, there will also be singularities at $t = 4x_1$, $t = 2x_2 + 2x_1$, etc., due to multiple reflections. The form of the reflection operator given in II, Eq. (3.11), shows that these reflections can be sorted out so that it can be determined whether a singularity in $u^r(t)$ is due to a multiple reflection or is the first reflection off of an interface, even if a first reflection and multiple reflection arrive simultaneously.

The problem of reconstructing $A(x)$ for $x > x_1$ can be approached in two ways. The first technique is to use $A(x)$ for $x < x_1$ and the data $u^i(t)$, $u^r(t)$ or $F(t)$, $G(t)$ to construct a set of data (u_x and u_t) at $x = x_1 -$. The method of Sec. 4 is then used to reconstruct $A(x)$ for $x_1 < x < x_2$. This process is continued, proceeding through the medium one layer at a time. The key points of this approach are now outlined.

To construct a set of data at $x = x_1$ in the electromagnetic case, Theorem 1 of I can be used. When specialized to the nonabsorptive case, that result states that for $0 < x < x_1$ and any t, the solution u of Eq. (2.5) is given by

$$\begin{aligned} u(x,t) = \tfrac{1}{2}\exp\left\{-\frac{1}{2}\int_0^x A(s)ds\right\}\{&(c_0+1) \\ &\times [u^i(x-t) + u^r(x+t)] \\ &-(c_0-1)[u^i(-x-t) + u^r(-x+t)] \\ &-\int_{-x}^{x} [u^i(y-t) + u^r(y+t)]N(x,y)\,dy\}. \end{aligned} \tag{6.1}$$

The function $N(x,y)$ in Eq. (6.1) is the solution of Eq. (4.4) and so all terms on the right-hand side of (6.1) are known. Thus, $u(x,t)$ is obtained by simple quadrature and, in particular, $u_x(x_1 - ,t)$ and $u_t(x_1,t)$ are also readily computed. Figure 3 illustrates this situation by showing the domain of dependence of $u(x_1 - ,t)$. To obtain a similar set of data along $x = x_1$ in the elastic case, it suffices to substitute Eq. (5.4) into Eq. (6.1).

In order to obtain a general result, assume now that data has been generated along the interface $x = x_k$; i.e., assume $u_x(x_k - ,t)$ and $u_t(x_k,t)$ are known. The process for obtaining $A(x)$ for $x_k < x < x_{k+1}$ is (see Fig. 4):

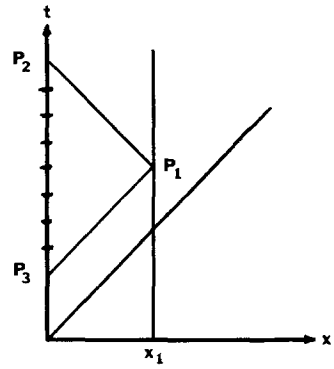

FIG. 3. Data along $x = 0$ and the solution N of Eq. (4.4) or (4.9) are used to compute u at P_1, where $P_1 = (x_1 - ,t)$. The data consists of the function $u^i(-t)$, $u^r(t)$ for t in the indicated interval, where P_2 and P_3 have coordinates $(0,t_0 + x_1)$, $(0,t_0 - x_1)$, respectively.

1. Define

$$u_k^i(-t') = -\frac{1}{2}\int_{x_k}^{t'+x_k} [u_x(x_k - ,s) - u_t(x_k,s)]\,ds, \tag{6.2a}$$

$$u_k^r(t') = \frac{1}{2}\int_{x_k}^{t'+x_k} [u_x(x_k - ,s) + u_t(x_k,s)]\,ds \quad \text{for } t' > 0. \tag{6.2b}$$

2. Set

$$c_k = c_k^-/c_k^+$$

and

$$f_k = (c_k - 1)/(c_k + 1)$$

and use Eq. (5.1) in the form

$$u_k^r(t') = f_k u_k^i(-t') + \int_0^{t'} R_k(s)u_k^i(s-t')\,ds \tag{6.3}$$

to obtain $R_k(t')$, $0 < t' < 2(x_{k+1} - x_k)$.

3. Solve the integral equation

$$\begin{aligned} N_k(x',y) = (c_k+1)R_k(x'+y) - f_k N_k(x',-y) \\ -\int_{-y}^{x'} R_k(y+s)N_k(x',s)\,ds \end{aligned} \tag{6.4}$$

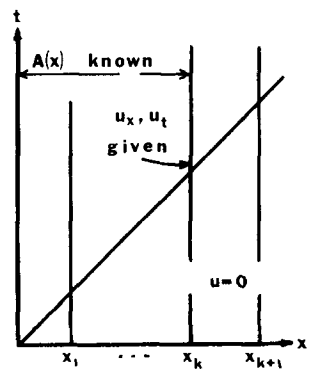

FIG. 4. This illustrates the first reconstruction process given in Sec. 6. The data u_x, u_t are given along $x = x_k -$ which allows a reflection kernel R_k to be constructed for the layer $x_k < x < x_{k+1}$. Equation (6.4) is then used to determine N_k, which in turn yields $A(x)$ for $x_k < x < x_{k+1}$.

for $N_k(x',y)$, $0<x'<(x_{k+1}-x_k)$, $-x'<y<x'$.

4. Solve for $A_k(x')$ via the analog of Eqs. (4.5) or (4.6). For example, from (4.5),

$$2N_k(x',x') = -A_k(0) - \tfrac{1}{4}(c_k+1)[2\{A_k(x')-A_k(0)\} + \int_0^{x'} A_k^2(s)\,ds].$$

5. Obtain $u_k(x',t')$ via the formula

$$u_k(x',t') = \tfrac{1}{2}\exp\left\{-\int_0^{x'} A_k(s)ds\right\} \times\{(c_k+1)[u_k^i(x'-t')+u_k^r(x'+t')] - (c_k-1)[u_k^i(-x'-t')+u_k^r(-x'+t')] - \int_{-x'}^{x'}[u^i(y-t')+u^r(y+t')]N_k(x',y)\,dy\}.$$

6. Translate back to (x,t) coordinates via

$$x = x_k + x',$$

$$t = x_k + t',$$

obtaining

$$A(x) = A(x_k+x') = A_k(x'), \quad 0<x'<(x_{k+1}-x_k)$$

and

$$u(x,t) = u(x_k+x', x_k+t') = u_k(x',t'), \quad 0<x'<(x_{k+1}-x_k),\ t'>0.$$

Thus, u_x and u_t can be determined along $x = x_{k+1}-$. Now steps 1–6 can be repeated to obtain $A(x)$ for $x_{k+1}<x<x_{k+2}$.

In determining $R_k(t')$ as given in Eq. (6.3), the methods of Sec. 5 can be used to convert (6.3) to a Volterra equation of the second kind, although some care must be exercised in doing so. For example, assume $u^i(t)$ or $F(t)$ is given by Eqs. (5.2) or (5.6). Then $u_k^i(-t')$ will contain δ-function singularities at $t'=0$ and also at points such as $t'=2x_1, t'=4x_1$, etc. [Examples of this are given by Eqs. (7.13) and (7.14).] These latter singularities are due to multiple reflections in the medium and their presence must be taken into account in performing the intergration in Eq. (6.3).

The alternate technique for reconstructing $A(x)$ is a straightforward specialization of the results of II, Lemma 3, to the case of a nonabsorptive medium (see Fig. 5). In contrast to the approach described above, data is not generated at the interfaces $x_1, x_2, \cdots$. Rather, the data at $x=0$ generates a function, denoted here by $\tilde{R}(t)$, via a generalization of Eq. (5.1). That generalization is found in II, Eq. (3.11). The function $\tilde{R}(t)$ is the reflection kernel for the composite medium, not just for a single layer as $R(t)$ or $R_k(t)$ are. This kernel is used to form an integral equation which is a generalization of Eq. (4.4), namely II, Lemma 3. This more general integral equation contains terms of the form $N(x,-y+k)$, where k is a known constant and $N(x,y)=0$ for $|y|>x$. The solution of the integral equation is then used to reconstruct $A(x)$ via II, Eq. (2.8).

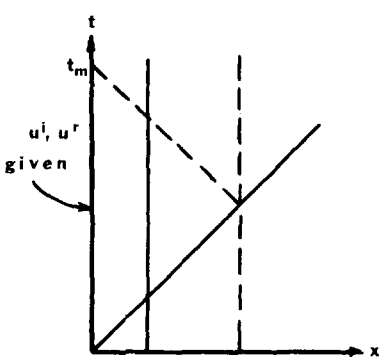

FIG. 5. The second reconstruction method of Sec. 6, which is a specialization of the results of II, uses a reflection kernel $\tilde{R}$ for the entire region $0<x<x_m$. The data $u^i(-t)$, $u^r(t)$ is given for $0<t<t_m=2x_m$.

This alternate technique has two advantages over the first technique.

1. It is not necessary to construct the data u_x, u_t along the interfaces $x_1, x_2, \ldots$.
2. In some cases it will be easier to construct $\tilde{R}(t)$ than $R_k(t)$. For example, if $u^i(t)=\delta(t)$, then in constructing $\tilde{R}(t)$ via II, Eq. (3.11), all multiple internal reflections have already been put in "by hand" throught the use of the shift operators introduced in that paper. Thus, the integral used in finding $\tilde{R}(t)$ is considerably simplified.

7. EXAMPLES

Example 1. The purpose of this first example is to demonstrate the inversion procedure for a single-layer elastic inverse problem with impedence mismatch between source and medium. Also, a technique for reducing Eq. (4.4) to the form of Eq. (4.9) will be shown.

Using the notation of Sec. 3, assume that

$$\rho c|_{0-} = 3$$

and

$$F(t) = \delta(t), \quad 0<t<2 \tag{7.1}$$

and that the resulting velocity is found to be

$$G(t) = -\tfrac{1}{2}\delta(t) + \tfrac{1}{2}e^{-t}, \quad 0<t<2. \tag{7.2}$$

The inversion procedure is as follows:

1. Comparing Eqs. (5.7) and (7.2), it follows that

$$c_0^+ = 2 \tag{7.3}$$

and therefore $c_0 = \frac{3}{2}$.

2. Equation (5.8) can be written as

$$\tfrac{4}{3}G_c(t) = -\tfrac{5}{6}R(t) + \int_0^t R(s)G_c(t-s)\,ds, \tag{7.4}$$

with

$$G_c(t) = \tfrac{1}{2}e^{-t}.$$

Upon differentiating (7.4) and using (7.4) to replace the integral, it follows that

$$R(t) = ke^{-0.4t} \tag{7.5}$$

for some constant k. Substituting (7.5) into (7.4) yields $k = -0.48$.

3. Substitute (7.5) into Eq. (4.4) and differentiate with respect to y. Using (4.4) to remove the integral term yields

$$5N_y(x,y) = N_y(x,-y) + 2N(x,-y) - 2N(x,y). \tag{7.6}$$

Set

$$N(x,y) = N_e(y) + N_o(y),$$

where (suppressing the x dependence)

$$N_e(y) = \tfrac{1}{2}[N(x,y) + N(x,-y)], \tag{7.7a}$$

$$N_o(y) = \tfrac{1}{2}[N(x,y) - N(x,-y)]. \tag{7.7b}$$

Substituting (7.7) into Eq. (7.6) and rearranging yields

$$3N_e'(y) + 2N_o(y) = -2N_o'(y). \tag{7.8}$$

Since the left- and right-hand sides of Eq. (7.8) are odd and even functions, respectively, it follows that both sides equal zero. Thus,

$$N_o(y) \equiv 0$$

and

$$N_e(y) \equiv c$$

for some constant c. Substituting $N(x,y) = c$ into (4.4) yields $c = -1$, so

$$N(x,y) \equiv -1.$$

4. Equation (4.6) now implies that

$$2A'(x) + A^2(x) = 0$$

with

$$A(0) = 2.$$

Thus,

$$A(x) = 2/(x+1).$$

5. From Eq. (4.12) it follows that $\rho c|_x = 2(x+1)^2$, $0 < x < 1$. This completes the solution. (See Fig. 6.)

It is possible to use an equation of the form of (4.9) instead of Eq. (4.4) even if $c_0 \neq 1$. This is done by constructing a set of data along $x = 0+$ and thereafter assuming $c_0 = 1$.

To construct data along $x = 0+$, note that

$$u_x(0+,t) = F(t)/c_0^+.$$

Now construct new "incident" and "reflected" fields according to

$$-u^i(-t) = \frac{1}{2}\int_0^t [u_x(0+,s) - u_t(0,s)]\,ds,$$

$$u^r(t) = \frac{1}{2}\int_0^t [u_x(0+,s) + u_t(0,s)]\,ds$$

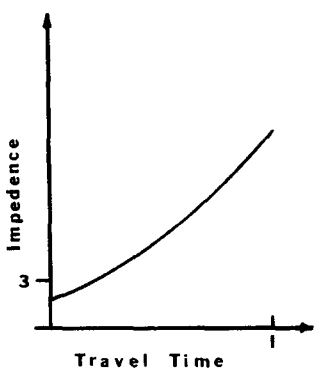

FIG. 6. Reconstructed impedence profile as a function of travel time for the impulse response given by Eq. (7.2).

or

$$-u^i(-t) = \frac{1}{2}\int_0^t \left[\frac{1}{c_0^+}F(s) - G(s)\right] ds, \tag{7.9a}$$

$$u^r(t) = \frac{1}{2}\int_0^t \left[\frac{1}{c_0^+}F(s) + G(s)\right] ds. \tag{7.9b}$$

Substituting (7.9) into Eq. (5.1) with $c_0 = 1$ yields

$$\int_0^t \left[\frac{1}{c_0^+}F(s) + G(s)\right] ds = -\int_0^t R(s)\left(\int_0^{t-s}\left[\frac{1}{c_0^+}F(y) - G(y)\right] dy\right) ds. \tag{7.10}$$

Using (7.1)–(7.3) in (7.10), differentiating twice, and solving for $R(t)$ as before yields

$$R(t) = \tfrac{1}{2}e^{-t/2}.$$

With this choice of $R(t)$, Eq. (4.9) can be solved in the same manner as Eq. (4.4), again yielding $N(x,y) \equiv -1$. The remainder of the solution proceeds as before. For this example, the solution technique based on Eq. (4.9) turns out to be no easier than that based on Eq. (4.4).

Clearly a similar technique can be used in the electromagnetic case to reduce Eq. (4.4) to Eq. (4.9).

Example 2. This example demonstrates the technique outlined in Sec. 6 for stepping through a multilayered medium. An electromagnetic inverse problem is considered with units chosen such that $c = 1/(\epsilon_0\mu_0)^{1/2} = 1$. Then, in the notation of Sec. 2,

$$u^i(t) = E^i(t),$$

$$u^r(t) = E^r(t).$$

Assume that the incident field is

$$u^i(-t) = \delta(t), \quad t < 14 \tag{7.11}$$

and the resulting reflected field is

$$u^r(t) = \Sigma a_n \delta(n-t), \quad t < 14, \tag{7.12}$$

where the summation is over $n = 0,6,8,10,12$ and $a_0 = -0.5$, $a_6 = 0.15$, $a_8 = -0.36$, $a_{10} = -0.036$, $a_{12} = 0.0114$. The presence of the term $\delta(6-t)$ in (7.12) implies that a discontinuity in permittivity occurs at $x = x_1 = 3$. Therefore, Eq. (5.1) takes the form

$$a_0\delta(-t) = f_0\delta(-t) + \int_0^t R(s)\delta(s-t)\,ds, \quad t < 6$$

and so $f_0 = -0.5$, $R(t) \equiv 0$, and $c_0 = \frac{1}{3}$. It follows that $N(x,y) \equiv 0$ and $A(x) \equiv 0$ for $0 < x < 3$. Then from Eq. (4.10),

$$z = x/3, \quad 0 < x < 3$$

and from (4.11),

$$\epsilon(z) = 9\epsilon_0, \quad 0 < z < 1.$$

Now Eq. (6.1) yields (for $x < 3$)

$$u(x,t) = \tfrac{1}{3}[-3a_0\delta(x-t) + 2\Sigma a_n\delta(n-x-t) + \Sigma a_n\delta(n+x-t)],$$

where the summation is over $n = 6,8,10,12$. Equation (6.2) then gives

$$u_1^i(-t') = \tfrac{1}{3}[-3a_0\delta(-t') + a_6\delta(6-t')], \quad t' < 8 \tag{7.13}$$

and

$$u_1^r(t') = \tfrac{2}{3}\Sigma a_n \delta(n-6-t'), \quad t'<8$$

summing over $n = 6,8,10,12$. The singularity in $u_1^r(t')$ at $t' = 2$ implies that $x_2 = x_1 + 1 = 4$. Equation (6.3) yields

$$\tfrac{2}{3}a_6\delta(-t') = -f_1 a_0 \delta(-t') - a_0 \int_0^{t'} R_1(s)\delta(s-t')\,ds, \quad t'<2.$$

Therefore, $f_1 = 0.2$, $R_1(t')\equiv 0$, $c_1 = 1.5$, and

$$\epsilon(z) = \epsilon(z_1 -)/c_1^2 = 4\epsilon_0, \quad 1<z<1.5.$$

Now for $3<x<4$ it follows that

$$\begin{aligned} u(x,t) = \tfrac{1}{12}[\,&\tfrac{36}{5}\delta(x-t) + 5a_6\delta(6+x-t) \\ &- a_6\delta(12-x-t) \\ &+ 10\Sigma a_n \delta(n-x-t) \\ &- 2\Sigma a_n \delta(n-6+x-t)], \end{aligned}$$

summing over $n = 8,10,12$. Then

$$u_2^i(-t') = \tfrac{1}{12}[\,\tfrac{36}{5}\delta(-t') - 2a_8\delta(2-t') - 2a_{10}\delta(4-t')], \quad t'<6 \tag{7.14}$$

and

$$u_2^r(t') = \tfrac{1}{12}[10\Sigma a_n \delta(n-8-t') - a_6\delta(4-t')], \quad t'<6$$

summing over $n = 8,10,12$. In this case the singularity in $u_2^r(t')$ at $t' = 2,4$ will be shown to be due to multiple reflections. To see this, note that Eq. (6.3) takes the form

$$u_2^r(t') = f_2 u_2^i(-t') + \int_0^{t'} R_2(s)u_2^i(s-t')\,ds$$

and is satisfied identically for $t'<6$ if $f_2 = -0.5$ and $R_2(t')\equiv 0$. Therefore, $c_2 = \frac{1}{3}$ and $\epsilon(z) = \epsilon(z_2 -)/c_2^2 = 36\epsilon_0$, $\frac{3}{2}<z<2$. The reconstructed profile is shown in Fig. 7. Since the reflected field was given only for $t<14$, nothing can be said about the profile for $z>2$.

If the alternate solution technique of Sec. 6 had been pursued, then the fact that $c_0 = c_2 = \frac{1}{3}$, $c_1 = 1.5$, $x_1 = 3$, $x_2 = 4$, $\tilde{R}(t)\equiv 0$ would follow easily from the form of the reflection operation given in II, Eq. (3.11); i.e., the results obtained above are immediately obvious by merely inspecting the coefficients a_0, a_6, a_8, a_{10}, a_{12}. The details are not pursued here because it would necessitate introducing a considerable amount of new notation.

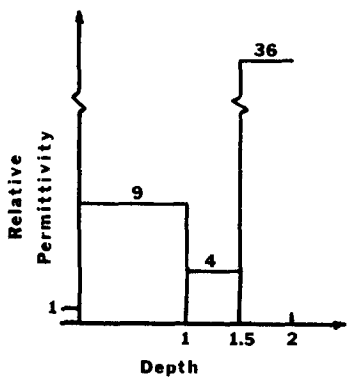

FIG. 7. Reconstructed permittivity profile as a function of depth for the incident and reflected fields given by Eqs. (7.11) and (7.12).

8. SUMMARY AND DISCUSSION

An algorithm for nonabsorbing electromagnetic and elastic inverse problems has been developed. In contrast to other approaches, the solution technique in this paper can be applied to media with discontinuous material properties. This also allows for an impedence mismatch between source and medium.

For a single layer medium with impedence mismatch at the source, the solution of the inverse problem is obtained by solving the integral equation (4.4) for $N(x,y)$. If there is no mismatch, then (4.4) reduces to the Marchenko equation (4.9). The function $R(t)$ in these equations is the reflection kernel obtained from the measured data. It is also shown via Example 1 of Sec. 7 that by altering the data, an equation of the form (4.9) can be used even when there is an impedence mismatch. However, in this case the function $R(t)$ is no longer the reflection kernel for the medium.

For a multilayered medium, two solution techniques are given. The first approach reconstructs the medium in the first layer and then uses the solution $N(x,y)$ of (4.4) or (4.9) to construct a new set of data at the interface between the first and second layers. Then the second layer is reconstructed, and the process continues. The technique is demonstrated in Example 2 of Sec. 7. The alternate technique for a multilayered medium does not reconstruct data at interfaces. It is a specialization of the results of II to a nonabsorbing medium, just as the approach in Sec. 4 is a specialization of the results in I. All of this work is carried out in the time domain.

Time-domain techniques for inverse problems have been examined carefully by Burridge.[7] In that paper, the classical results of Gel'fand–Levitan[10] and Marchenko[8] are shown to have a relatively simple derivation when cast in the time domain. Burridge also claims that the results of Gopinath and Sondhi[5,6] are applicable to the elastic inverse problem for the case in which the impedence has jump discontinuities. Although Gopinath and Sondhi do not seem to have realized this when first developing their approach, it appears that the results of their experiment on a discontinuous vocal tract[6] could have been sharpened if the presence of additional δ-function singularities had been detected. Recent time domain work by Symes[11] shows the well-posed nature of the elastic inverse problem for the case of continuous (actually H^1) impedence.

Inverse problems in the frequency domain have received more attention than those in the time domain. Among authors who have considered continuous (as opposed to discretized) inverse problems, Moses and deRidder[12] and Kay and Moses[13] adapted the Gel'fand–Levitan technique to the electromagnetic case, and Ware and Aki[14] did the same for elastic media. In each case the appropriate model equation, (2.1) or (3.1), was mapped into a Schrödinger equation by transformation of dependent and independent variables. These transformations disallow jump discontinuities in the material properties. This has also been pointed out by Berryman and Greene,[15] who go on to develop a discretized ap-

proach to the inverse problem for elastic media with discontinuities. The frequency domain technique of Razavy[16] does appear to be capable of handling discontinuous coefficients. Although it seems the technique is not immediately applicable to elastic media [because the term $E_x\psi_x$ is eliminated from his Eq. (1)], the model equation used is well suited to the electromagnetic case.

ACKNOWLEDGMENTS

The author thanks Professor Brian DeFacio and Professor James Corones for helpful discussions, and the referee for a useful comment. This work was supported by the U.S. D.O.E., Contract No. W-7405-eng-82, Division of Engineering, Mathematical and Geosciences, budget code AK-01-04-02.

[1]R. J. Krueger, "An inverse problem for a dissipative hyperbolic equation with discontinuous coefficients," Quart. Appl. Math. **34**, 129 (1976).
[2]R. J. Krueger, "An inverse problem for an absorbing medium with multiple discontinuities," Quart. Appl. Math. **36**, 235 (1978).
[3]V. H. Weston, "On the inverse problem for a hyperbolic dispersive partial differential equation," J. Math. Phys. **13**, 1952 (1972).
[4]R. J. Krueger, "Numerical aspects of a dissipative inverse problem," IEEE Trans. Antenna Propag. **AP-29**, 253 (1981).
[5]B. Gopinath and M. M. Sondhi, "Inversion of the telegraph equation and the synthesis of nonuniform lines," Proc. IEEE **59**, 383 (1971).
[6]M. M. Sondhi and B. Gopinath, "Determination of vocal tract shape from impulse response at the lips," J. Acoust. Soc. Am. **49**, 1867 (1971).
[7]R. Burridge, "The Gel'fand–Levitan, the Marchenko, and the Gopinath–Sondhi integral equations of inverse scattering theory, regarded in the context of inverse impulse-response problems," Wave Motion **2**, 305 (1980).
[8]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[9]H. Hochstadt, *Integral Equations* (Wiley-Interscience, New York, 1973), p. 31.
[10]I. M. Gel'fand and B. M. Levitan, "On the determination of a differential equation from its spectral function," Izv. Akad. Nauk. SSSR **15**, 309 (1951) [Amer. Math. Soc. Transl. (2) **1**, 253 (1955)].
[11]W. W. Symes, "Impedence profile inversion via the first transport equation," preprint.
[12]H. E. Moses and C. M. deRidder, "Properties of dielectrics from reflection coefficients in one dimension," MIT Lincoln Lab. Tech. Report 322 (1963).
[13]I. Kay and H. E. Moses, "Reflectionless transmission through dielectrics and scattering potentials," J. Appl. Phys. **27**, 1503 (1956), Appendix 2.
[14]J. A. Ware and K. Aki, "Continuous and discrete inverse-scattering problems in a statified elastic medium. I. Plane waves at normal incidence," J. Acoust. Soc. Am. **45**, 911 (1969).
[15]J. G. Berryman and R. R. Greene, "Discrete inverse methods for elastic waves in layered media," Geophys. **45**, 213 (1980).
[16]M. Razavy, "Determination of the wave velocity in an inhomogeneous medium from the reflection coefficient," J. Acoust. Soc. Am. **58**, 956 (1975).

Feynman path integral and Poisson processes with piecewise classical paths

Ph. Combe,[a] R. Hoegh-Krohn,[b] R. Rodriguez,[a] M. Sirugue,[c] and M. Sirugue-Collin[d]
C.N.R.S.-Luminy-Case 907, Centre de Physique Théorique, F-13288, Marseille, Cedex 2, France

(Received 19 November 1980; accepted for publication 5 December 1980)

We prove the existence of a Feynman integral formula for gentle perturbations of the harmonic oscillator. This result is extended to a Bose relativistic theory.

PACS numbers: 03.65.Db, 03.65.Fd

1. INTRODUCTION

In a previous paper, hereafter quoted as I[1] (see also Ref. 2), we have shown that it is possible to write a Feynman formula for the evolution in time of a quantum state for a large class of evolution operators. The general result can be applied to many quantum systems and in particular to some gentle perturbations of Hamiltonians of systems with n degrees of freedom whose spectrum is purely discrete. As a special case it is possible to apply this result to some perturbations of the harmonic oscillator. Unfortunately, then, the role of the classical trajectories is not as clear as previously.

For this reason in this paper we derive another expression for the Feynman path integral which is more adapted to the harmonic oscillator case. This problem has been already considered in the framework of Fresnel integrals.[3] However, our present treatment is quite different in spirit since it is based on the notion of Poisson measures whose role has been already emphasized in Ref. 4.

The "free evolutions" we consider are actually the quasifree evolutions we considered in Ref. 5, more precisely, quantum systems whose classical phase spaces have a group structure and which are quantized according to the Weyl procedure. In that case any one-parameter group of automorphisms of the phase space group gives rise to a one-parameter group of automorphisms of the quantum algebra naturally associated.

We perturb these free evolutions by an interaction which is of the type described in I. Then we can define generalized Poisson processes whose paths are the classical paths except for a finite number of jumps. These Poisson processes allow to write the time evolution with respect to the perturbed evolution of the state vector, as an integral on a natural measure space of a function of the process. The function of the process is the natural analog of the classical free action of the problem. This is essentially the content of Sec. 2.

In Sec. 3 we specialize in the case where the classical phase space is $\mathbb{R}^{2n}$, and to those potentials which are multiplicative in the "x" representation. The free evolutions that we consider are given by a one-parameter subgroup of the metaplectic group, and we give the relevant formulas for its typical one-parameter subgroups.

In Sec. 4 we extend the previous results to an infinite number of degrees of freedom. These results naturally apply to some important model of field theory like the sine–Gordon model, but we postpone to a forthcoming paper a discussion on this point.

2. GENERAL CASE

Let us consider an abelian locally compact group G and ζ a continuous multiplier on G, namely a continuous function from $G \times G$ to the torus, such that

$$\zeta(g_1,g_2)\zeta(g_1g_2,g_3) = \zeta(g_1,g_2g_3)\zeta(g_2,g_3) \quad \forall g_1,g_2,g_3 \in G \tag{2.1}$$

In the following, to simplify the formulas, we shall assume that $\zeta(g,g^{-1}) = 1$. A Weyl system on G, with respect to this multiplier, is a mapping of G into the unitary group of a C^*-algebra $\overline{\Delta(G,\zeta)}$ such that

$$\begin{aligned} &\text{(i)} W_g W_{g'} = \zeta(g,g') W_{gg'}, \\ &\text{(ii)} W_g^* = W_{g^{-1}}, \\ &\text{(iii)} \Delta(G,\zeta) = \left\{ \sum_{i=1}^{n} \lambda_i W_{g_i};\ \lambda_i \in \mathbb{C}, g_i \in G \right\} \end{aligned} \tag{2.2}$$

is uniformly dense in $\overline{\Delta(G,\zeta)}$.

$\overline{\Delta(G,\zeta)}$ is unique up to isomorphisms if and only if ζ is nondegenerate in the sense that

$$G_\zeta = \{ g \in G; b_\zeta(g,g') = 1 \ \forall g' \in G \} = \{e\}, \tag{2.3}$$

where b_ζ is the antisymmetric bicharacter on G associated with ζ, namely

$$b_\zeta(g,g') = \zeta(g,g')\,\overline{\zeta(g',g)} \tag{2.4}$$

(see, e.g., Ref. 6).

Let $t \to \alpha_t$ be a one-parameter group of automorphisms of G such that

$$\zeta(\alpha_t(g),\alpha_t(g')) = \zeta(g,g') \quad \forall g,g' \in G. \tag{2.5}$$

We can define a family of $*$-automorphisms of $\overline{\Delta(G,\zeta)}$ according to

$$\tilde{\alpha}_t(W_g) = \lambda(t,g) W_{\alpha_t^{-1}(g)} \quad \forall g \in G, \tag{2.6}$$

where λ, is an application from $R \times G$ to the torus. In many cases λ is 1, and in the following we shall mostly consider this case. But for the sake of completeness we can give an example where λ is nontrivial and nevertheless (2.6) defines a group of $*$-automorphisms. Let G be $\mathbb{R}^2$ with the usual multiplier (in what follows $\hbar = 1$)

$$\zeta((q,p),(q',p')) = \exp[(i/2)(pq' - p'q)]; \tag{2.7}$$

[a] Centre de Physique Théorique, CNRS Marseille, and Faculté des Sciences de Luminy.
[b] Mathematiske Institute, Universiteit i Oslo, and Centre de Physique Théorique, CNRS Marseille.
[c] Centre de Physique Théorique, CNRS Marseille.
[d] Centre de Physique Théorique, CNRS Marseille, and Université de Provence.

 0022-2488/82/030405-07$02.50 © 1982 American Institute of Physics

we can take

$$\alpha_t(q,p) = (q + pt, p) \tag{2.8}$$

and

$$\lambda(t,(q,p)) = \exp[ia(pt^2/2 - qt)] \tag{2.9}$$

for any real a.

It is obvious that the necessary and sufficient conditions for (2.6) to define a one-parameter group of $*$-automorphisms of $\overline{\Delta(G,\zeta)}$ are

$$\begin{aligned} &\lambda(t + t',g) = \lambda(t,g)\lambda(t',\alpha_{-t}(g)), \\ &\lambda(t,gg') = \lambda(t,g)\lambda(t,g'), \\ &\lambda(0,g) = \lambda(t,e) = 1. \end{aligned} \tag{2.10}$$

Let us consider a representation of $\overline{\Delta(G,\zeta)}$ where a one-parameter group of $*$-automorphisms α_t is unitarily implemented by a (continuous) unitary group $\mathcal{U}_t = \exp(-iH_0 t)$. Let us consider a "potential" V which is of the type

$$V = \int_G d\mu(g) W_g, \tag{2.11}$$

where μ is a bounded measure on G. One can write the operator $\exp(iTH_0)\exp[-iT(H_0 + V)]$ as a convergent Dyson's expansion

$$\begin{aligned} &\exp(iTH_0)\exp[-iT(H_0 + V)]\psi \\ &\quad = \sum_{n>0}(-i)^n \int_0^T dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 \\ &\times \int_G \cdots \int_G d\mu(g_n)\cdots d\mu(g_1) e^{it_nH_0} W_{g_n} e^{-it_nH_0} \cdots e^{it_1H_0} W_{g_1} e^{-it_1H_0}\psi \end{aligned} \tag{2.12}$$

or, using the fact that $\exp(itH_0)$ implements α_t,

$$\begin{aligned} &\exp(iTH_0)\exp[-iT(H_0 + V)]\psi \\ &\quad = \sum_{n>0}(-i)^n \int_0^T dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 \\ &\times \int_G \cdots \int_G d\mu(g_n)\cdots d\mu(g_1) W_{\alpha_{-t_n}(g_n)} \cdots W_{\alpha_{-t_1}(g_1)}\psi. \end{aligned} \tag{2.13}$$

Notice that even if $\mathcal{U}_t$ is not continuous (i.e., if H_0 does not exist), the right-hand side of (2.13) defines a solution of the equation

$$\begin{aligned} &i\frac{d}{dt}\psi(t) = \mathcal{U}_t^{-1} V \mathcal{U}_t \psi(t), \\ &\psi(0) = \psi, \end{aligned} \tag{2.14}$$

which is precisely the equation we want to analyze in the following.

Using the Weyl relation (2.2) and the fact that ζ is α_t invariant, we can rewrite the expression (2.13)

$$\begin{aligned} &\exp(iTH_0)\exp[-iT(H_0 + V)]\psi \\ &\quad = \sum_{n>0}(-i)^n \int_0^T dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 \\ &\times \int_G \cdots \int_G d\mu(g_n)\cdots d\mu(g_1) \\ &\times \prod_{i=1}^{n} \overline{\zeta(\alpha_{t_{i-1}-t_n}(g_n)\cdots\alpha_{t_i-t_n}(g_i) g_{i-1}, g_{i-1}^{-1})} \\ &\times W_{\alpha_{-t_n}(g_n)\cdots\alpha_{-t_1}(g_1)}\psi. \end{aligned} \tag{2.15}$$

Finally let us rewrite the matrix elements of (2.15) between states of the form $\psi = W_g\Phi, g\in G$ where Φ is the cyclic vector of the representation.

$$\begin{aligned} &(W_g\Phi \,|\exp(iTH_0)\exp[-iT(H_0 + V)]\, W_g\Phi) \\ &\quad = \sum_{n>0}(-i)^n \int_0^T dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 \\ &\times \int_G \cdots \int_G d\mu(g_n)\cdots d\mu(g_1) \\ &\times \prod_{i=1}^{n} \overline{\zeta(\alpha_{t_{i-1}-t_n}(g_n)\cdots\alpha_{t_i-t_n}(g_i) g_{i-1}, g_{i-1}^{-1})} \\ &\times b_\zeta(\alpha_{-t_n}(g_n)\cdots\alpha_{-t_1}(g_1), g)(\Phi \,|\, W_{\alpha_{-t_n}(g_n)\cdots\alpha_{-t_1}(g_1)}\Phi). \end{aligned} \tag{2.16}$$

It is that last expression that we want to analyze in terms of Poisson processes. First we shall describe a realization of the underlying measure space Ω (see, e.g., Refs. 1, 2, 4, and also 7, 8.)

Let Ω be the measure space

$$\Omega = \bigcup_{n>0} \Omega_n, \tag{2.17}$$

where Ω_n is the set

$$\Omega_n = \{\eta = (n,t_i,g_i); 0 < t_1 < t_2 \cdots < t_n < T; g_i \in G\}. \tag{2.18}$$

Let $\mathcal{F}$ be the smallest σ-Borel algebra on Ω which contains the open sets

$$\vartheta^{(n)}_{\{a_i\}\{\mathcal{B}_i\}} = \{(n,t_i,g_i); t_i \in a_i, g_i \in \mathcal{B}_i\}, \tag{2.19}$$

where the $\mathcal{B}_i$ are Borel subsets of G and the a_i are measurable ordered subsets in $[0,T]$ in the sense that

$$t_i \in a_i, \quad t_j \in a_j, \quad i < j \rightarrow t_i < t_j. \tag{2.20}$$

There exists on Ω a unique bounded measure P_μ associated with a bounded measure μ on G such that

$$P_\mu(\vartheta^{(n)}_{\{a_i\}\{\mathcal{B}_i\}}) = \prod_{i=1}^{n} |a_i| \mu(\mathcal{B}_i), \tag{2.21}$$

where $|a_i|$ is the Lebesgue measure of a_i. P_μ is the Poisson measure induced by μ.

It is natural to consider a Poisson process G_τ based on $[0,T]$ with values in G:

$$G_\tau(n,t_i,g_i) = \begin{cases} e, & \tau \in]t_n, T], \\ \alpha_{\tau - t_n}(g_n^{-1}), & \tau \in]t_{n-1}, t_n], \\ \alpha_{\tau - t_i}(g_i^{-1})\cdots\alpha_{\tau - t_n}(g_n^{-1}), & \tau \in]t_{i-1}, t_i], \\ \alpha_{\tau - t_1}(g_1^{-1})\cdots\alpha_{\tau - t_n}(g_n^{-1}), & \tau \in [0, t_1] \end{cases} \tag{2.22}$$

This process has "classical paths" except for a finite number of jumps. We also define the increment of the process over the jump

$$\Delta G_\tau(n,t_i,g_i) = \lim_{\epsilon \to 0} G_{\tau-\epsilon}\, G_{\tau+\epsilon}^{-1}. \tag{2.23}$$

Evidently

$$\Delta G_{t_i}(n,t_k,g_k) = g_i \tag{2.24}$$

Otherwise

$$\Delta G_\tau(n,t_k,g_k) = e.$$

For this process we can note that there is a natural decreasing adapted family $\mathcal{F}_t$, $t\in[0,T]$, of Borel σ-algebras, namely, for each t the smallest Borel σ-algebras containing the open subsets

$$\vartheta^{(n)}_{t\{a_i\}|\mathscr{H}_t|} = \bigcup_{k>0}\bigcup_{b_i\subset[0,t]} \vartheta^{(n)}_{\{b_i\}|\{a_i\},G\cdots G|\mathscr{H}_t|}. \tag{2.25}$$

Moreover, the functionals

$$\eta\in\Omega \to Z(G^{-1},\Delta G)(\eta) = \prod_{i=1}^{n} \zeta(G_{t_i}(\eta)^{-1},\Delta G_{t_i}(\eta)), \tag{2.26}$$

$$\eta\in\Omega \to (\Phi \mid W_{G_0^{-1}}\Phi)(\eta) = (\Phi \mid W_{G_0^{-1}(\eta)}\Phi), \tag{2.27}$$

and

$$\eta\in\Omega \to B_g(G_0)(\eta) = b_\zeta(g,G_0^{-1}(\eta)) \tag{2.28}$$

are bounded measurable functionals as well as the functional

$$\eta\in\Omega\to(-i)^{N_0(\eta)}, \tag{2.29}$$

where N_τ is the usual Poisson process

$$\eta\in\Omega\to N_t(\eta)\in\mathbf{Z}, \tag{2.30}$$

where

$$N_t(n,t_i,g_i) = \sum_{i=1}^{n} \theta_+(t_i - t). \tag{2.31}$$

θ_+ is the step function.

If we gather all these results, we can rewrite the matrix elements (2.16)

$$(W_g\Phi \mid \exp(iTH_0)\exp[-iT(H_0+V)]W_g\Phi) = \int_\Omega P(d\eta)\{(-i)^{N_0}\overline{Z(G^{-1},\Delta G)}B_g(G_0)(\Phi \mid W_{G_0^{-1}}\Phi)\}(\eta). \tag{2.32}$$

The case where λ is nontrivial can be handled quite easily. The next section is devoted to more specific examples.

3. MULTIPLICATIVE POTENTIALS IN QUANTUM MECHANICS

In this section we specialize ourselves to quantum mechanics with one degree of freedom, so that $G = \mathbf{R}^2$ with the multiplier ζ defined in (2.7). Consequenlty, $W_g = W_{qp}$, $(q,p)\in\mathbf{R}^2$, and

$$W_{qp} = \exp[i(qP - pQ)], \tag{3.1}$$

where P and Q are the usual momentum and position operators.

We restrict ourselves to potentials which are multiplicative in the "x" representation, viz., of the form

$$V = \int_{\mathbf{R}} d\mu(p) W_{0p}. \tag{3.2}$$

In that case the Poisson measure is supported by the set of elements $\eta\in\Omega$ of the form

$$\eta = (n,t_i,(0,p_i)). \tag{3.3}$$

Hence we can restrict ourselves to a new measure space we still denote by Ω, constructed on R in the same way as previously.

On $\mathbf{R}^2$ there are basically three nontrivial kinds of automorphisms which are of the following type[5]:

$$\alpha^1_t(q,p) = (q+pt,p),$$
$$\alpha^2_t(q,p) = (\cos(\omega t)q + (1/\omega)\sin(\omega t)p, -\omega\sin(\omega t)q + \cos(\omega t)p), \tag{3.4}$$
$$\alpha^3_t(q,p) = (\cosh(\omega t)q + (1/\omega)\sinh(\omega t)p, \omega\sinh(\omega t)q + \cosh h(\omega t)p)$$

for ω a real number. Correspondingly, it is convenient to define the three processes:

$$Q^1_t(n,t_i,p_i) = \sum_{j=1}^{n} p_j(t_j - t)\theta_+(t_j - t),$$
$$Q^2_t(n,t_i,p_i) = \sum_{j=1}^{n} \frac{p_j}{\omega}\sin[\omega(t_j - t)]\theta_+(t_j - t), \tag{3.5}$$
$$Q^3_t(n,t_i,p_i) = \sum_{j=1}^{n} \frac{p_j}{\omega}\sinh[\omega(t_j - t)]\theta_+(t_j - t)$$

with continuous paths. Except for a finite number of times, the trajectories are the trajectories of the free motion.

These processes are the projections of the process defined in (2.22):

$$(Q^1,P^1)_t(n,t_i,(0,p_i)) = \left(\sum_{j=1}^{n} p_j(t_j - t)\theta_+(t_j - t), \sum_{j=1}^{n} p_j\theta_+(t_j - t)\right)$$
$$(Q^2,P^2)_t(n,t_i,(0,p_i)) = \left(\sum_{j=1}^{n} \frac{p_j}{\omega}\sin[\omega(t_j - t)]\theta_+(t_j - t), \sum_{j=1}^{n} p_j\cos[\omega(t_j - t)]\theta_+(t_j - t)\right), \tag{3.6}$$
$$(Q^3,P^3)_t(n,t_i,(0,p_i)) = \left(\sum_{j=1}^{n} \frac{p_j}{\omega}\sinh[\omega(t_j - t)]\theta_+(t_j - t), \sum_{j=1}^{n} p_j\cosh[\omega(t_j - t)]\theta_+(t_j - t)\right).$$

In terms of these processes one can rewrite the function $\overline{Z(G^{-1},\Delta G)}$, and, moreover, taking into account that ζ is an antisymmetric bicharacter, one has

$$\overline{Z((Q^1,P^1)^{-1},\Delta(Q^1,P^1))}(n,t_i,(0,p_i)) = \exp\left[\tfrac{1}{2}i\sum_{\substack{r,s=1\\ r>s}}^{n} p_r p_s(t_r - t_s)\right] \tag{3.7}$$

which for convenience we denote by

$$\exp\left(\tfrac{1}{2}i\int_0^T Q^1_\tau \dot{Q}^1_\tau d\tau\right)(n,t_i,p_i).$$

$$\overline{Z((Q^2,P^2)^{-1},\Delta(Q^2,P^2))}(n,t_i,(0,p_i)) = \exp\left(\tfrac{1}{2}i\sum_{\substack{r,s=1\\ r>s}}^{n} \frac{1}{\omega}\sin[\omega(t_r - t_s)]p_r p_s\right) \tag{3.8}$$

which again for notational convenience we rewrite

$$\exp\left[\tfrac{1}{2}i\int_0^T Q^2_\tau\left(\frac{\partial^2}{\partial t^2} + \omega^2\right)Q^2_\tau d\tau\right](n,t_i,p_i).$$

Finally

$$\overline{Z((Q^3,P^3)^{-1},\Delta(Q^3,P^3))}(n,t_i,(0,p_i))$$
$$= \exp\left\{\tfrac{1}{2}i\sum_{\substack{r,s=1\\ r>s}}^{n}\frac{1}{\omega}\sinh[\omega(t_r-t_s)]p_rp_s\right\} \tag{3.9}$$

which we denote by

$$\exp\left[\tfrac{1}{2}i\int_0^T Q^3_\tau\left(\frac{\partial^2}{\partial t^2}-\omega^2\right)Q^3_\tau d\tau\right](n,t_i,p_i).$$

It is probably more suggestive to rewrite the processes as

$$(Q^i,P^i)(n,t_i,(0,p_i)) = (Q^i,\dot{Q}^i)(n,t_i,p_i) \tag{3.10}$$

and

$$S^1_0(Q^1)(n,t_i,p_i) = \tfrac{1}{2}\sum_{\substack{r,s=1\\ r>s}}^{n}p_rp_s(t_r-t_s), \tag{3.11}$$

$$S^2_0(Q^2)(n,t_i,p_i) = \tfrac{1}{2}\sum_{\substack{r,s=1\\ r>s}}^{n}\frac{p_rp_s}{\omega}\sin[\omega(t_r-t_s)], \tag{3.12}$$

$$S^3_0(Q^3)(n,t_i,p_i) = \tfrac{1}{2}\sum_{\substack{r,s=1\\ r>s}}^{n}\frac{p_rp_s}{\omega}\sinh[\omega(t_r-t_s)]. \tag{3.13}$$

It is quite obvious to calculate the functions B_ζ in our cases

$$B_\zeta((q,p),(Q^1,P^1)_0)(n,t_i,(0,p_i))$$
$$= \exp\left[i\left(\sum_{r=1}^{n}t_rp_rp-\sum_{r=1}^{n}p_rq\right)\right],$$
$$B_\zeta((q,p),(Q^2P^2)_0)(n,t_i,(0,p_i))$$
$$= \exp\left\{i\sum_{r=1}^{n}\left[\frac{p_r}{\omega}\sin(\omega t_r)p-p_r\cos(\omega t_r)q\right]\right\}, \tag{3.14}$$
$$B_\zeta(q,p),(Q^3P^3)_0)(n,t_i,(0,p_i))$$
$$= \exp\left\{i\sum_{r=1}^{n}\left[\frac{p_r}{\omega}\sinh(\omega t_r)p-p_r\cosh(\omega t_r)q\right]\right\}.$$

We can rewrite these functions altogether under the form

$$B_\zeta((q,p),(Q^j\dot{Q}^j)_0)(n,t_i,(0,p_i)) = \exp[i(pQ^j_0-q\dot{Q}^j_0)](n,t_i,p_i). \tag{3.15}$$

Consequently, if we gather these results, we have

$$(W_{qp}\Phi\,|\exp(iTH_0)\exp[-iT(H_0+V)]\,W_{qp}\Phi)$$
$$= \int_\Omega P(d\eta)[e^{-i\pi N_0/2}e^{iS_0(Q)}e^{-ipQ_0+iq\dot{Q}_0}$$
$$\times(\Phi W_{-Q_0,-\dot{Q}_0)}\Phi)](\eta). \tag{3.16}$$

It is interesting to notice that in the example which has been given where λ is nontrivial one can incorporate it in S_0:

$$S'_0(Q^1) = S_0(Q^1)+a\int_0^T Q^1_\tau d\tau,$$

at least formally, which is to be interpreted as

$$S'_0(Q^1)(n,t_i,p_i) = \sum_{\substack{r,s=1\\ r>s}}^{n}p_rp_s(t_r-t_s)+a\sum_{r=1}^{n}\frac{p_rt_r^2}{2}\ .$$

4. BOSE QUANTUM FIELD

In the general case of a neutral scalar Bose field we recall elementary facts concerning Klein–Gordon equation in order to make contact between the usual formalism and the one we shall need later.

Let us consider the Klein–Gordon equation

$$\left(\frac{\partial^2}{\partial t^2}-\sum_{i=1}^{s}\frac{\partial^2}{\partial x_i^2}+m^2\right)\varphi(x,t)=0. \tag{4.1}$$

It is equivalent to the system

$$\frac{\partial}{\partial t}\varphi(x,t)=\pi(x,t),$$
$$\frac{\partial}{\partial t}\pi(x,t)=\left(\sum_{i=1}^{s}\frac{\partial^2}{\partial x_i^2}-m^2\right)\varphi(x,t), \tag{4.2}$$

which can be rewritten through Fourier transformation:

$$\tilde{f}(k)=\frac{1}{(2\pi)^{s/2}}\int_{\mathbf{R}^s}f(x)e^{ikx}dx,$$
$$\frac{\partial}{\partial t}\tilde{\varphi}(k,t)=\tilde{\pi}(k,t), \tag{4.3}$$
$$\frac{\partial}{\partial t}\tilde{\pi}(k,t)=-\omega(k)^2\tilde{\varphi}(k,t),$$

where $\omega(k)=(k^2+m^2)^{1/2}$, but in general we can consider even functions of k, strictly positive and indefinitly differentiable. Let h be the set of infinitely differentiable functions from $\mathbb{R}^s$ to $\mathbb{C}$, with rapid decrease at infinity and satisfying

$$f(k)=\overline{f(-k)}. \tag{4.4}$$

It is a real vector space. Moreover, $\mathscr{H}=h\oplus h$ is a symplectic space with respect to the following symplectic form σ (i.e., real bilinear, antisymmetric, nondegenerated form):

$$\sigma((f,g),(f',g'))$$
$$= \int_{\mathbf{R}^s}dk\,[\,\overline{f'(k)}g(k)-f(k)\,\overline{g'(k)}\,]. \tag{4.5}$$

We can rewrite the Klein-Gordon equation on $\mathscr{H}$:

$$\frac{\partial}{\partial t}(f,g)=K(f,g), \tag{4.6}$$

where K is the real linear operator:

$$K(f,g)=(f',g'),$$
$$f'(k)=g(k),\quad g'(k)=-\omega(k)^2f(k). \tag{4.7}$$

K has the following "polar decomposition":

$$K=J\Omega=\Omega J, \tag{4.8}$$

where $\Omega(f,g)=(f',g')$,

$$f'(k)=\omega(k)f(k),\quad g'(k)=\omega(k)g(k) \tag{4.9}$$

and

$$J(f,g)=(f',g'),$$
$$f'(k)=\omega(k)^{-1}g(k),\quad g'(k)=-\omega(k)f(k). \tag{4.10}$$

J is a complex structure on $\mathscr{H}$, viz., ,

$$J^2=-1 \tag{4.11}$$

and

$$\sigma((f,g),J(f,g))$$
$$= \int_{\mathbf{R}^s}dk\,[\omega(k)|f(k)|^2+\omega(k)^{-1}|g(k)|^2]\geqslant 0. \tag{4.12}$$

Moreover,

$$\sigma(J(f,g),J(f',g'))=\sigma((f,g),(f',g')). \tag{4.13}$$

K leaves $\mathscr{H}$ invariant and one can exponentiate K to find in

$\mathscr{H}$ the solution of the Klein–Gordon equation

$$\exp(KT)(f,g) = (f_T,g_T),$$

$$f_T(k) = \cos[\omega(k)T]f(k) + \omega(k)^{-1}\sin[\omega(k)T]g(k), \quad (4.14)$$

$$g_T(k) = -\omega(k)\sin[\omega(k)T]f(k) + \cos[\omega(k)T]g(k);$$

$\exp(KT)$ leaves σ invariant.

J as a complex structure defines a bijection Γ_J of $\mathscr{H}$ onto a complex pre-Hilbert space $\Gamma_J(\mathscr{H})$ with scalar product

$$\langle\Gamma_J(f,g)|\Gamma_J(f',g')\rangle = \sigma((f,g),J(f',g')) + i\sigma((f,g),(f',g')); \quad (4.15)$$

moreover,

$$\Gamma_J J = i\Gamma_J, \quad (4.16)$$

$$\sigma((f,g),(f',g')) = \mathrm{Im}\langle\Gamma_J(f,g)|\Gamma_J(f',g')\rangle. \quad (4.17)$$

Consequently, if we consider $\overline{\Gamma_J(\mathscr{H})}$ completion of $\Gamma_J(\mathscr{H})$ with respect to this scalar product, $\overline{\Gamma_J(\mathscr{H})}$ is the space of pairs of classes of functions $(f,g):\mathbb{R}^s\to\mathbb{C}^2$ such that

$$\text{(i)}\quad f(k) = \overline{f(-k)},$$

$$\text{(ii)}\quad g(k) = \overline{g(-k)}, \quad (4.18)$$

$$\text{(iii)}\quad \int_{\mathbb{R}^s} dk\,[\omega(k)|f(k)|^2 + \omega(k)^{-1}|g(k)|^2] < \infty.$$

According to (4.17), σ can be extended to that space into a real bilinear nondegenerate antisymmetric form which we denote by the same symbol. A Weyl system $\overline{\Gamma_J(\mathscr{H})}$, taken as a real symplectic space $\overline{\mathscr{H}}$ with respect to σ, in a mapping $(f,g)\to W_{fg}$ into the unitary group of an Hilbert space such that

$$W_{fg}^* = W_{(-f,-g)}, \quad (4.19)$$

$$W_{fg}W_{f'g'} = \exp[\tfrac{1}{2}i\sigma((f,g),(f',g'))]\,W_{f+f',g+g'}. \quad (4.20)$$

If both $\lambda\to W_{\lambda f,0}$ and $\mu\to W_{0,\mu g}$ are continuous, then

$$W_{fg} = \exp[i(\Pi(f) - \Phi(g))], \quad (4.21)$$

where the self-adjoint operators $\Pi(f)$ and $\Phi(g)$ satisfy

$$[\Phi(g),\Pi(f)] \subseteq i\int_{\mathbb{R}^s} \overline{g(k)}f(k)dk. \quad (4.22)$$

Let $\overline{\Delta(\overline{\mathscr{H}},\sigma)}$ be the C^*-algebra generated by the Weyl operators. This algebra has a lot of states, and we shall be interested mainly by the quasifree states which are those states ω which satisfy (see Ref. 9 and the references therein):

$$\omega(W_{fg}) = \exp[-\tfrac{1}{4}s((f,g),(f,g))], \quad (4.23)$$

where s is a real, bilinear, symmetric, positive form and satisfies

$$|\sigma((f,g),(f',g'))|^2 \leqslant s((f,g),(f,g))s((f',g'),(f',g'))$$
$$f',g' \in \overline{\mathscr{H}}. \quad (4.24)$$

In fact, we are interested by the states which are invariant both by translations and by the evolution, where translation is defined as the group of *-automorphisms of $\overline{\Delta(\mathscr{H},\sigma)}$, which satisfies

$$\tau_a(W_{fg}) = W_{f_a,g_a}, \quad a\in\mathbb{R}^s, \quad (4.25)$$

where

$$f_a(k) = e^{ika}f(k) \quad (4.26)$$

and the evolution is given by

$$\alpha_t(W_{fg}) = W_{(fg)_{-t}}, \quad (4.27)$$

where $(f,g)\to(fg)_t$ is the extension of $\exp(Kt)$ to $\overline{\mathscr{H}}$. For quasifree states which are invariant both w.r.t translations and α_t we have the following theorem (cf. Ref. 3, Theorem 8.1)

Theorem (4.28): Any quasifree state of $\overline{\Delta(\mathscr{H},\sigma)}$ which is invariant w.r.t translation and α_t is given by a real bilinear positive form on $\overline{\mathscr{H}}$ such that

$$s((f,g),(f,g)) = \sigma((f,g),D(f,g)), \quad (4.28)$$

where

$$D:\overline{\mathscr{H}}\to\overline{\mathscr{H}}, \quad D(f,g) = (f',g')$$

with

$$f'(k) = a(k)f(k) + b(k)g(k),$$

$$g'(k) = -\omega(k)^2 b(k)f(k) + a(k)g(k).$$

a and ωb are L_∞ functions satisfying

$$a(k) = -\overline{a(k)} = -a(-k),$$

$$b(k) = \overline{b(k)} = b(-k),$$

and

$$\omega(k)b(k) - ia(k) \geqslant 1, \quad \omega(k)b(k) \geqslant 1.$$

For the sake of completeness we sketch the proof which is slightly different from the one in Ref. 3:

Let V be the operator from $\overline{\Gamma_J(\mathscr{H})}$ to $L_2(\mathbb{R}^s,dk)$

$$V(f,g) = \omega(k)^{1/2}f(k) - i\omega(k)^{-1/2}g(k). \quad (4.29)$$

It satisfies

$$V\Gamma_J J = iV\Gamma_J \quad (4.30)$$

and

$$\|V\Gamma_J(f,g)\|_{L_2} = \|\Gamma_J(f,g)\|_{\overline{\Gamma_J(\mathscr{H})}}. \quad (4.31)$$

Hence $\overline{\Gamma_J(\mathscr{H})}$ is isomorphic to $L^2(\mathbb{R}^s,dk)$. Any symmetric, bilinear, J-invariant form on $\overline{\mathscr{H}}$ is transported into a Hermitian form on $L^2(\mathbb{R}^s,dk)$, and, since it is everywhere defined,

$$s((f,g),(f,g)) = (V\Gamma_J(f,g)|AV\Gamma_J(f,g)), \quad (f,g)\in\overline{\mathscr{H}}, \quad (4.32)$$

where A is a bounded operator greater than 1.

Translation invariance implies that there exists an L_∞ function $A(k)$ such that[10]

$$(Af)(k) = A(k)f(k), \quad f\in L_2(\mathbb{R}^s,dk),$$

which, moreover, is greater or equal to 1.

On the other hand, one can explicitly invert V, viz.,

$$V^{-1}(f) = \Gamma_J(f',g'), \quad f\in L_2(\mathbb{R}^s,dk), \quad (4.33)$$

where

$$f'(k) = \tfrac{1}{2}\omega(k)^{-1/2}[f(k) + \overline{f(-k)}],$$

$$g'(k) = \tfrac{1}{2}i\omega(k)^{1/2}[f(k) - \overline{f(-k)}]. \quad (4.34)$$

Hence $V^{-1}AV$ defined from $\overline{\Gamma_J(\mathscr{H})}$ to $\overline{\Gamma_J(\mathscr{H})}$ is given in matrix form

$$V^{-1}AV=\begin{vmatrix}\alpha & \beta\\ \gamma & \delta\end{vmatrix}, \tag{4.35}$$

where

$$\alpha(k)=\delta(k)=\tfrac{1}{2}[A(k)+A(-k)],$$
$$\omega(k)\beta(k)=-\omega(k)^{-1}\gamma(k)=-\tfrac{1}{2}i[A(k)-A(-k)].$$

Finally, if we recall that

$$\langle\Gamma_J(f,g)|\Gamma_J(f,g)\rangle=\sigma((f,g),J(f,g)), \tag{4.36}$$

one has

$$D=J\Gamma_J^{-1}V^{-1}AV\Gamma_J=\begin{vmatrix}a & b\\ c & d\end{vmatrix}, \tag{4.37}$$

with

$$a(k)=d(k)=\tfrac{1}{2}i[A(k)-A(-k)],$$
$$\omega(k)b(k)=-\omega(k)^{-1}c(k)=\tfrac{1}{2}[A(k)+A(-k)].$$

Hence the necessary conditions

$$a(k)=\overline{-a(k)}=-a(-k),$$
$$b(k)=\overline{b(k)}=b(-k),$$
$$\omega(k)b(k)-ia(k)\geqslant 1 \quad \text{or} \quad \omega(k)b(k)\geqslant\tfrac{1}{2}.$$

Conversely, to any pair of L_∞ functions a and b satisfying the previous conditions there corresponds an invariant quasifree state of $\overline{\Delta(\mathscr{H},\sigma)}$.

A particulary important example is the Fock state which corresponds to

$$b(k)=\omega(k)^{-1},\quad a(k)=0, \tag{4.38}$$

for which

$$\omega_F(W_{fg})=\exp\left\{-\frac{1}{4}\int dk\,[\omega(k)|f(k)|^2+\omega(k)^{-1}|g(k)|^2]\right\}. \tag{4.39}$$

It should be stressed that the previous theorem depends strongly on the choice of $\overline{\mathscr{H}}$ to construct the Weyl algebra. For instance, the following quasifree state,

$$\omega_\mu(W_{fg})=\exp\left\{-\frac{1}{4}s((f,g),(f,g))-\int d\mu(k)[\omega(k)|f(k)|^2+\omega(k)^{-1}|g(k)|^2]\right\}, \tag{4.40}$$

$(f,g)\in\mathscr{H}$, where μ is a positive bounded measure, is both translation invariant and invariant with respect to the evolution. These states have at least an extension to $\overline{\Delta(\overline{\mathscr{H}},\sigma)}$ since $(\overline{\Delta\mathscr{H}},\sigma)$ is a subalgebra of $\overline{\Delta(\overline{\mathscr{H}},\sigma)}$. Some of them at least are not factorial since if μ is a Dirac measure on $k=0$ we have explicitly

$$\exp\left[-\frac{1}{4}s((f,g),(f,g))-\lambda_0^2(\omega(0)|f(0)|^2+\omega(0)^{-1}|g(0)|^2)\right]=\int_0^\infty\int_0^{2\pi}dr\,\frac{d\theta}{2\pi}\,\frac{r}{\lambda_0}\,e^{-r^2/2\lambda_0}\omega_{\theta r}(W_{fg}), \tag{4.41}$$

where

$$\omega_{\theta r}(W_{fg})=\exp\{-\tfrac{1}{4}s((f,g),(f,g))+ir(\omega(0)^{1/2}(f/0)+\overline{f(0)})\cos\theta+\omega(0)^{-1/2}[g(0)+\overline{g(0)}]\sin\theta\}.$$

In (4.40) one can, for instance, even choose μ as an unbounded positive measure and takes 0 for the state and for those $\omega^{1/2}f$ or $\omega^{-1/2}g$, which are not μ-square integrable.

Now let us come back to our problem, which is to treat perturbations of the free evolution given by potentials of the form

$$V=\int d\mu(g)W_{0g}. \tag{4.42}$$

We can apply the machinery of the previous section, at least if the underlying group $\overline{\mathscr{H}}$ is endowed with the discrete topology. This is not a too serious restriction since some physically relevant interaction, viz., the exponential interaction with cutoffs, falls in this class. Instead we can take as basic group $\mathscr{S}(\mathbb{R}^3)\oplus\mathscr{S}(\mathbb{R}^3)$ and use the fact that it is a nuclear space. We write a Feynman formula for the matrix elements of the perturbations of the free evolution with the state ψ corresponding to a quasifree state as we have previously described,

$$(\psi|e^{iH_0T}e^{-iT(H_0+V)}\psi)=\int_\Omega P(d\eta)\left\{e^{-i\pi N(0)/2}\times\exp\left[-(i/2)\int_0^T dt\int_\lambda dx\,X(x,t)\{\Box+m^2\}X(x,t)\right]\right\}(\eta) \tag{4.43}$$

for a suitable process X we shall describe later, and a suitable definition of the measurable functional

$$\eta\to\exp\left[-(i/2)\int_0^T dt\int dx\,X(x,t)\{\Box+m^2\}X(x,t)\right](\eta).$$

The process we shall describe depends explicitly on the state we choose. But first of all let us introduce a process which will always occur.

Let $(\Omega,\mathscr{F},P)$ be the measure space defined as previously except that now $G\equiv\bar{h}$ and P is the Poisson measure constructed from the bounded measure μ associated with the potential (4.42)

Let us consider the following process:

$$X_R(k,t)(n,t_i,f_i)=\sum_{i=1}^n\Delta_R(k,t_i-t)f_i(k), \tag{4.44}$$

where

$$\Delta_R(k,t)=\frac{1}{(k^2+m^2)^{1/2}}\sin[(k^2+m^2)^{1/2}t]\theta_+(t). \tag{4.45}$$

In x space

$$X_R(x,t)(n,t_i,f_i)=\frac{1}{(2\pi)^{s/2}}\int dk\,e^{ikx}X_R(k,t)(n,t_i,f_i)=\sum_{i=1}^n\int dy\,\Delta_R(x-y,t_i-t)f_i(y). \tag{4.46}$$

In order to show (4.43), one has to note that

$$\sigma((0,g_r)_{t_r-t_s},(0,g_s)) = \int \frac{dk}{(k^2+m^2)^{1/2}} \sin[(k^2+m^2)^{1/2}(t_r-t_s)]g_r(k)\,\overline{g_s(k)}; \tag{4.47}$$

hence

$$\sum_{r>s}\sigma((0,g_r)_{t_r-t_s},(0,g_s)) = \sum_{r,s}\int dk\ \overline{g_r(k)}\,\Delta_R(k,t_r-t_s)g_s(k). \tag{4.48}$$

The process $X_R(x,t)$ verifies an inhomogeneous Klein–Gordon equation, namely,

$$\{\Box + m^2\}X_R(x,t)(\eta) = \sum_{i=1}^{n} g_i(x)\delta(t-t_i), \tag{4.49}$$

where $\eta = (n,t_i,g_i)\in\Omega$. More generally, for a quasifree state as described in Theorem (4.28), given by the two L_∞ functions a, ωb, we can consider an adapted process:

Let us consider the function $\Delta^{a,b}(k,t)$ (cf. Ref. 3):

$$\Delta^{ab}(k,t) = b(k)\cos[\omega(k)t] - \frac{1}{\omega(k)}a(k)\sin[\omega(k)t] - \frac{i}{\omega(k)}\sin[\omega(k)|t|], \tag{4.50}$$

and the corresponding process

$$X^{ab}(x,t)(\eta) = \sum_{i=1}^{n}\Delta^{ab}*g_i(x,t-t_i). \tag{4.51}$$

This process satisfies an equation similar to that of X_R namely

$$\{\Box + m^2\}X^{ab}(x,t)(\eta) = -2i\sum_{i=1}^{n} g_i(x)\delta(t-t_i). \tag{4.52}$$

Let Φ be the cyclic vector corresponding to the quasifree state through the GNS construction. One can compute explicitly the functional

$$\eta\in\Omega\to(W_{f_0g_0}\Phi\,|\,W_{-X_R(\cdot,0),-\dot{X}_R(\cdot,0)}W_{f_0g_0}\Phi)(\eta)$$

corresponding to a dense set of vectors in the representation space $\mathscr{H}$ corresponding to the state we have considered:

$$\begin{aligned}(W_{f_0g_0}\Phi\,|\,W_{-X_R(\cdot,0),-\dot{X}_R(\cdot,0)}W_{f_0g_0}\Phi)(\eta) &= \exp\Big(-\frac{1}{4}\sum_{i,j}\int dk\,\{b(k)\cos[\omega(k)(t_i-t_j)] + \frac{a(k)}{\omega(k)}\sin[\omega(k)(t_i-t_j)]\}\,\overline{g_i(k)}g_j(k)\Big)\\ &\times\exp\Big\{i\sum_j\int \overline{f_0(k)}\cos[\omega(k)t_j]g_j(k) + \overline{g_0(k)}\,\frac{\sin[\omega(k)t_j]}{\omega(k)}g_j(k)\Big\}.\end{aligned} \tag{4.53}$$

We can gather this formula with formula (4.48) and rewrite:

$$\begin{aligned}Z(X)(W_{f_0g_0}\Phi\,|\,W_{-X_R(\cdot,0),-\dot{X}_R(\cdot,0)}W_{f_0g_0}\Phi)(\eta) &= \exp\Big[-\frac{1}{4}\sum_{i,j}\int dk\,\Delta^{ab}(k,t_i-t_j)\,\overline{g_i(k)}g_j(k)\Big]\\ &\times\exp[i\sigma((f_0,g_0),(X_R(\cdot,0),\dot{X}_R(\cdot,0)))],\end{aligned} \tag{4.54}$$

which we rewrite

$$\begin{aligned}Z(X)(W_{f_0g_0}\Phi\,|\,W_{-X_R(\cdot,0),-\dot{X}_R(\cdot,0)}W_{f_0g_0}\Phi)(\eta) &= \exp\Big[-\tfrac{1}{2}i\int_0^T dt\int dx\,X^{ab}(x,t)\{\Box+m^2\}X^{ab}(x,t)\Big]\\ &\times\exp\Big\{-i\int dx\,[\overline{f_0(x)}\dot{X}_R(x,0) - \overline{g_0(x)}\,X_R(x,0)]\Big\}(\eta).\end{aligned} \tag{4.55}$$

Consequently, we can state the following proposition:

Proposition (4.56): Let ω^{ab} be a quasifree state of $\Delta(\overline{\mathscr{H}},\sigma)$ which is translation invariant and invariant with respect to the free evolution. Let H_0 be the Hamiltonian in the corresponding representation. Let V be a perturbation given by a bounded measure $V = \int d\mu(g)W_{0g}$. Let Φ^{ab} be the cyclic vector corresponding to ω^{ab}, and $f_0,g_0\in\overline{\mathscr{H}}$; then

$$\begin{aligned}&(W_{f_0g_0}\Phi^{ab}\,|\exp(iTH_0)\exp[-iT(H_0+V)]\,W_{f_0g_0}\Phi^{ab})\\ &= \int_\Omega P(d\eta)\Big(\exp\Big(-\tfrac{1}{4}i\int_0^T dt\int dx X^{ab}(x,t)\{\Box+m^2\}X^{ab}(x,t)\\ &\quad\times\exp[-\tfrac{1}{2}i\pi N(0)]\exp\Big\{i\int dx\Big[\overline{g_0(x)}\,X_R(x,0) - \overline{f_0(x)}\dot{X}_R(x,0)\Big]\Big\}\Big)(\eta),\end{aligned} \tag{4.56}$$

where X^{ab} and X_R are the processes which have been previously defined.

This proposition naturally applies to the sine–Gordon model with ultraviolet cutoffs and more generally to exponential interactions, but we shall discuss those models in more detail along the previous lines in another paper.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor S. Albeverio, Professor Ph. Blanchard, and Professor G. Rideau for discussions. Two of us, Professor R. H. K. and M. S., acknowledge Professor L. Streit for his kind hospitality in Bielefeld, where some part of this work was done, and the V.S.P Mathematiesierung of the University of Bielefeld for financial support.

[1] Ph. Combe, R. Hoegh-Krohn, R. Rodriguez, M. Sirugue, and M. Sirugue-Collin, "Poisson Processes on Groups and Feynman Path Integrals," Commun. Math. Phys. 77, 269 (1980).
[2] Ph. Combe, R. Hoegh-Krohn, P. Rodriguez, M. Sirugue, and M. Sirugue-Collin, in *Proceedings of Louvain Conference on Functional Integration*, edited by J.-P. Antoine and E. Tirapegui (Plenum, New York, 1980).
[3] R. Hoegh-Krohn and S. Albeverio, *Mathematical Theory of Feynman Path Integral*, Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1976).
[4] A. M. Chebotarev and V. P. Maslov, in *Proceedings of the Colloquium on Feynman Path Integral*, Marseille 1978, Springer Lecture Notes in Physics (Springer-Verlag, Berlin, 1979).
[5] Ph. Combe, R. Rodriguez, M. Sirugue, and M. Sirugue-Collin in *Proceedings of the Fifth International Colloquium on Group Theoretical Methods in Physics, Montreal*, edited by R. T. Sharp and B. Coleman (Academic, New York, 1979).
[6] J. Manuceau, M. Sirugue, D. Testard, and A. Verbeure, Commun. Math. Phys. **32**, 231 (1973).
[7] D. Shale, J. Funct. Anal. **33**, 1 (1979).
[8] H. Heyer, *Probability Measure on Locally Compact Groups* (Springer-Verlag, Berlin, 1977).
[9] D. Testard, Thesis (Marseille University, 1970).
[10] L. Hormander, Acta Math. **104**, 93 (1960).

Theoretical basis for Coulomb matrix elements in the oscillator representation

Augustine C. Chen
Physics Department, St. John's University, Jamaica, New York 11439

(Received 1 September 1981; accepted for publication 16 October 1981)

Hydrogenic wave functions in the spherical and parabolic bases are shown to correspond, respectively, to a restricted set of wave functions of a four-dimensional harmonic oscillator and its coupled pair of two-dimensional oscillators. This correspondence provides the theoretical basis for algebraic calculations of Coulomb matrix elements in the oscillator representation.

PACS numbers: 03.65.Ge

1. INTRODUCTION

The connection between the hydrogen atom and the harmonic oscillator has been investigated by many authors. Schrödinger[1] showed by his method of double factorization that radial eigenfunctions of a four-dimensional oscillator are "closely connected, in some cases identical, with the radial eigenfunctions of the Kepler motion" in coordinate space. Using group theoretical methods, Chaćon, Levi, and Moshinsky[2] identified the radial eigenstates of the "pseudo-Coulomb" problem[3] with the radial eigenstates of the four-dimensional oscillator. Čížek and Paldus[4] introduced a transformation relating the radial Schrödinger equations of the two systems. More recently, Barut, Schneider, and Wilson[5] connected the Kepler motion in three dimensions with oscillatory motion in four dimensions by the Kustaanheimo–Stiefel transformation. Utilizing explicitly two sets of polar coordinates in perpendicular spaces, Chen[6] showed the correspondence between the two systems and expanded the total hydrogenic wave functions as a linear combination of four-dimensional oscillator wave functions in Hermite polynomials. On the other hand, Englefield[7] and others[8] related, by comparison, the two separated equations for the hydrogen atom in three-dimensional parabolic coordinates with the equation of a two-dimensional oscillator and thus the atom is tacitly taken to be equivalent to a pair of uncoupled two-dimensional oscillators.

In this paper, we reexamine the correspondence between the two systems, hydrogen atom and the four-dimensional oscillator. In so doing, we extend Schrödinger's original conclusion on radial eigenfunctions and supplement Chen's earlier work by explicitly identifying the quantum numbers of the corresponding states. We will also take a closer look at the separation of the hydrogenic equation in parabolic coordinates with regard to its invertibility. This subtle point has never been thoroughly examined.[9] Because of its elusiveness, it has been inadvertently overlooked or, perhaps, ignored. In Sec. 2, we examine the correspondence. In Sec. 3, we investigate the separation procedure. After having established the theoretical foundation we discuss, in Sec. 4, the algebraic calculation of Coulomb matrix elements in the oscillator representation. Many of the mathematical expressions in the paper are familiar to the readers. They are needed because of the subtlety of the point at hand. Its exposition calls for a detailed comparison of the relevant mathematical expressions.

2. THE CORRESPONDENCE

The dynamical group of the hydrogen atom[10] is the conformal group SO(4,2). It contains SO(4) and SO(2,1) as subgroups. The dynamical group of a four-dimensional harmonic oscillator[11] is the symplectic group Sp(8). It contains the direct product $\mathrm{Sp}(2)\times\mathrm{SO}(4)$ as a subgroup. The groups SO(2,1) and Sp(2) are isomorphic. Therefore, the two systems have in common the same underlying symmetry. It is natural to expect that their correspondence can be established in a four-dimensional coordinate space. Fock[12] showed that the bound-state wave functions of the hydrogen atom in momentum space are related to the angular part of the four-dimensional spherical harmonics. The trick used was a stereographic projection of the momentum space onto a unit sphere in four-dimensional Euclidean space. We will show that the position-space wave functions of the atom in the spherical and parabolic bases are related to the wave functions of a four-dimensional oscillator and its coupled pair of two-dimensional oscillators, respectively. The trick now, as we shall see, is a superposition of two phase angles of the pair of two-dimensional oscillators to form the single azimuthal angle of the three-dimensional atom.

We begin with the Schrödinger equation, in atomic units, of a four-dimensional oscillator with frequency ω,

$$-\frac{1}{2}\sum_{i=1}^{4}\left(\frac{\partial^2}{\partial s_i^2}-\omega^2 s_i^2\right)\psi = N\omega\psi. \tag{1}$$

Letting

$$y_i = (\omega)^{1/2} s_i, \tag{2}$$

we can rewrite Eq. (1) as

$$-\frac{1}{2}\sum_{i=1}^{4}\left(\frac{\partial^2}{\partial y_i^2}-y_i^2\right)\psi = N\psi. \tag{3}$$

Now we introduce the following coordinates, which consist of two sets of polar coordinates in perpendicular spaces,

$$y_1 = y\cos\alpha\cos\beta, \tag{4a}$$

$$y_2 = y\cos\alpha\sin\beta, \tag{4b}$$

$$y_3 = y\sin\alpha\cos\gamma, \tag{4c}$$

$$y_4 = y\sin\alpha\sin\gamma. \tag{4d}$$

The Kustaanheimo–Stiefel transformation relating these coordinates to the three-dimensional Cartesian coordinates is shown in the Appendix. In these coordinates, Eq. (3) takes

 0022-2488/82/030412-05$02.50 © 1982 American Institute of Physics

the form

$$H\psi = -\frac{1}{2}\left(\frac{\partial^2}{\partial y^2} + \frac{3\partial}{y\partial y} + \frac{4\mathbf{K}^2}{y^2} - y^2\right)\psi = N\psi. \quad (5)$$

Equation (5) is invariant under rotations in four dimensions. The six generators for the SO(4) group are[13]:

$$K_z = -\frac{1}{2}i\left(\frac{\partial}{\partial\beta} + \frac{\partial}{\partial\gamma}\right), \quad (6)$$

$$K_\pm = \tfrac{1}{2}\exp[\pm i(\beta+\gamma)] \times\left(\mp\frac{\partial}{\partial\alpha} + i\tan\alpha\frac{\partial}{\partial\beta} - i\cot\alpha\frac{\partial}{\partial\gamma}\right), \quad (7)$$

$$\mathbf{N}_z = -\frac{1}{2}i\left(\frac{\partial}{\partial\beta} - \frac{\partial}{\partial\gamma}\right), \quad (8)$$

$$\mathbf{N}_\pm = \tfrac{1}{2}\exp[\pm i(\beta-\gamma)] \times\left(\pm\frac{\partial}{\partial\alpha} + i\tan\alpha\frac{\partial}{\partial\beta} + i\cot\alpha\frac{\partial}{\partial\gamma}\right). \quad (9)$$

The significance of these operators are further discussed in the Appendix. The Casimir operator of the group $4\mathbf{K}^2$ in Eq. (5) is given by

$$4\mathbf{K}^2 = 4\mathbf{N}^2 = -\frac{\partial^2}{\partial\alpha^2} - \sec^2\alpha\frac{\partial^2}{\partial\beta^2} - \csc^2\alpha\frac{\partial^2}{\partial\gamma^2} + (\tan\alpha - \cot\alpha)\frac{\partial}{\partial\alpha}. \quad (10)$$

Its eigenvalue is denoted by $k(k+2)$. The solutions of Eq. (5) have the form

$$\psi_{Nkm_1m_2} = R_{Nk}(y^2)Y_{km_1m_2}(\alpha\beta\gamma), \quad (11)$$

where

$$R_{Nk}(y^2) = \text{const}\times\exp(-\tfrac{1}{2}y^2)\,y^k L^{k+1}_{\frac{1}{2}(N-k)-1}(y^2), \quad (12)$$

$$Y_{km_1m_2}(\alpha\beta\gamma) = d^{\frac{1}{2}k}_{m_1m_2}(2\alpha)\exp[im_1(\beta+\gamma)]\exp[im_2(\beta-\gamma)] \quad (13)$$

are the radial and angular parts of the wave function given, respectively, in terms of the Laguerre functions and the three-dimensional reduced rotation matrices.

To establish the correspondence between the hydrogen atom and the oscillator, we set[6] $\omega = (-8E)^{1/2} = 2/n$, $N\omega = 4$, $y^2 = \omega r$, $2\alpha = \theta$, and $\beta+\gamma = \varphi$ (or $\beta-\gamma = \varphi$ as discussed in the Appendix). It then follows that $N = 2n$ and $4\mathbf{K}^2 = 4\mathbf{L}^2$, i.e., $k = 2l$ since $k(k+2) = 4l(l+1)$. We also note

$$d^l_{mo}(\theta)\exp(im\varphi) = \left(\frac{4\pi}{2l+1}\right)^{1/2}Y_{lm}(\theta,\varphi). \quad (14)$$

Therefore, the four-dimensional oscillator states specified by the quantum numbers $|N = 2n, k = 2l, m_1 = m, m_2 = 0\rangle$ correspond to hydrogenic states $|nlm\rangle$. Thus we have extended Schrödinger's result on radial eigenfunctions by completely identifying the corresponding states of the two systems.

The generators of the group Sp(2)~SO(2,1) are given by

$$T_1 = \frac{1}{4}\left(-\frac{\partial^2}{\partial y^2} - \frac{3\partial}{y\partial y} + \frac{4\mathbf{K}^2}{y^2} - y^2\right), \quad (15)$$

$$T_2 = -\frac{1}{2}i\left(y\frac{\partial}{\partial y} + 2\right), \quad (16)$$

$$T_3 = \frac{1}{4}\left(-\frac{\partial^2}{\partial y^2} - \frac{3\partial}{y\partial y} + \frac{4\mathbf{K}^2}{y^2} + y^2\right) = \tfrac{1}{2}H. \quad (17)$$

They satisfy the commutation relations

$$[T_1,T_2] = -iT_3, \quad [T_3,T_1] = iT_2, \quad [T_2,T_3] = iT_1. \quad (18)$$

The radial wave functions of the two systems satisfy the following:

$$T_3R_{ab} = bR_{ab}, \quad (19)$$

$$T_\pm R_{ab} = [(b\pm a)(b\mp a\pm 1)]^{1/2}R_{ab\pm 1}, \quad (20)$$

where

$$T_\pm = T_1 \pm iT_2 \quad (21)$$

and the indices are given by $a = \frac{1}{2}(k+2)$, $b = \frac{1}{2}N$, and $a = l+1$, $b = n$, respectively, for the oscillator and the atom. The respective normalized radial wave functions are

$$R_{Nk} = \omega\left[\frac{2(\frac{1}{2}N - \frac{1}{2}k - 1)!}{(\frac{1}{2}N + \frac{1}{2}k)!}\right]^{1/2} \times\exp(-\tfrac{1}{2}y^2)\,y^k L^{k+1}_{\frac{1}{2}(N-k)-1}(y^2), \quad (22)$$

$$R_n = \frac{2}{n^2}\left[\frac{(n-l-1)!}{(n+l)!}\right]^{1/2}\exp(-\tfrac{1}{2}\rho)\,\rho^l L^{2l+1}_{n-l-1}(\rho), \quad (23)$$

where $y^2 = \omega s^2$ and $\rho = 2r/n$. The volume elements for the normalization integrals are, respectively, $d\tau = s^3ds$ and $dV = r^2dr$. An exact correspondence requires renormalizing R_{Nk} by using $d\tau' = s^5ds$ instead of $d\tau$. Consequently, we have

$$R_{nl} = (\tfrac{1}{2}\omega/n)^{1/2}R_{Nk} = R_{Nk}/n. \quad (24)$$

This relationship makes possible the algebraic calculation of the mean values of r^ρ in the oscillator representation in Sec. 4.

3. THE SEPARATION PROCEDURE

By letting

$$\mu = \left(\frac{\omega}{2}\right)^{1/2}u = y\cos\alpha, \quad (25a)$$

$$\nu = \left(\frac{\omega}{2}\right)^{1/2}v = y\sin\alpha, \quad (25b)$$

we can separate Eq. (3) or (5) into the following pair of equations:

$$-\frac{1}{2}\left(\frac{\partial^2}{\partial\mu^2} + \frac{1}{\mu}\frac{\partial}{\partial\mu} + \frac{1}{\mu^2}\frac{\partial^2}{\partial\beta^2} - \mu^2\right) = N_1\psi_1, \quad (26)$$

$$-\frac{1}{2}\left(\frac{\partial^2}{\partial\nu^2} + \frac{1}{\nu}\frac{\partial}{\partial\nu} + \frac{1}{\nu^2}\frac{\partial^2}{^2\partial\gamma^2} - \nu^2\right) = N_2\psi_2, \quad (27)$$

where $N_1 + N_2 = N$. The total wave function of the system, $\psi_{N_1N_2m_1m_2}$, is given by the product $\psi_1\psi_2$. We are interested in solutions for which $\beta+\gamma = \varphi$, $N_1+N_2 = 2n$, $m_1 = m$, and $m_2 = 0$. These are the solutions that give rise to the hydrogenic wave functions in the parabolic basis. Let us denote these solutions $\psi_{N_1N_2m0}$ by $|N_1N_2m;\omega\rangle$. They are given by

$$|N_1N_2m;\omega\rangle = R(\mu^2)R(\nu^2)\exp[im(\beta+\gamma)]$$
$$= \frac{\omega}{4\pi}\left[\frac{(\frac{1}{2}N_1-\frac{1}{2}m-\frac{1}{2})!(\frac{1}{2}N_2-\frac{1}{2}m-\frac{1}{2})!}{(\frac{1}{2}N_1+\frac{1}{2}m-\frac{1}{2})!(\frac{1}{2}N_2+\frac{1}{2}m-\frac{1}{2})!}\right]^{1/2}$$
$$\times\exp[im(\beta+\gamma)](\mu\nu)^m\exp\left(-\frac{\mu^2+\nu^2}{2}\right)$$
$$\times L^{m}_{(\frac{1}{2})(N_1-m-1)}(\mu^2)L^{m}_{(\frac{1}{2})(N_2-m-1)}(\nu^2). \quad (28)$$

We now decompose each of the operators in Eq. (19) into two pieces, T_i^1 and T_i^2, one acting on $R(\mu^2)$ and the other on $R(\nu^2)$. Then we have

$$T_3^j R_{ab} = bR_{ab}, \quad (29)$$

$$T_\pm^j R_{ab} = [(b\pm a)(b\mp a\pm 1)]^{1/2}R_{ab\pm 1}, \quad (30)$$

where $a=\frac{1}{2}(m+1)$ and $b=\frac{1}{2}N_j, j=1,2$. The volume element used for the normalization is $d\tau = 4uvdudvd\beta d\gamma$ with the range of β and γ being 0 to 2π. The volume element used in normalizing hydrogenic states in the parabolic basis, given by

$$|n_1n_2m\rangle = \left[\frac{n_1!n_2!}{n^4\pi(n_1+m)!(n_2+m)!}\right]^{1/2}\exp[\frac{1}{2}(\xi+\eta)/n]$$
$$\times\exp(im\varphi)(\xi\eta/n^2)^m L^{m}_{n_1}(\xi/n)L^{m}_{n_2}(\eta/n), \quad (31)$$

is $dV=\frac{1}{4}(\xi+\eta)d\xi d\eta d\varphi$. As before, we need to renormalize $|N_1N_2m;\omega\rangle$ with respect to the volume element $d\tau' = (u^2+v^2)uvdudvd\beta d\gamma$. As a result, we have

$$|n_1n_2m\rangle = \left(\frac{\omega}{n}\right)^{1/2}S\left(\frac{2}{n\omega}\right)\Big|N_1N_2m;\omega\Big\rangle$$
$$= \frac{(2)^{1/2}}{n}\Big|N_1N_2m\Big\rangle, \quad (32)$$

since $\omega = 2/n$ for the hydrogen atom. The scaling operator is given by

$$S(\lambda) = -iT_2\ln\lambda. \quad (33)$$

The two sets of quantum numbers are related as follows: $n_1=\frac{1}{2}(N_1-m-1)$, $n_2=\frac{1}{2}(N_2-m-1)$. We will use Eq. (32) to calculate a transition matrix in Sec. 4.

Now we come to the subtle point regarding the invertibility of our separation procedure which reflects the underlying symmetry of the problem and the noninvertibility of the separation procedure using parabolic coordinates. It is clear that each of the two equations (26) and (27), complete with its own angular dependence, represents a *bona fide* two-dimensional system, namely, a two-dimensional oscillator. The corresponding equations in parabolic coordinates, however, do not have the proper angular dependence. The separated equations given in the literature are of the form[14]

$$-\left[\frac{d}{d\xi}\left(\xi\frac{d}{d\xi}\right)-\frac{m^2}{4\xi}-\frac{E\xi}{2}\right]R(\xi) = Z_1R(\xi), \quad (34)$$

$$-\left[\frac{d}{d\eta}\left(\eta\frac{d}{d\eta}\right)-\frac{m^2}{4\eta}-\frac{E\eta}{2}\right]R(\eta) = Z_2R(\eta), \quad (35)$$

where Z_1 and Z_2 are the separation constants and the product $R(\xi)R(\eta)$ gives the "radial" part of Eq. (31). It is well known that Eqs. (34) and (35) obtain *after* the angular dependent term in the unseparated equation is replaced by $-m^2$ upon operating on the angular dependent part of the wave function. If the angular dependence is *given back* to the separated equations by writing $\partial^2/\partial\varphi^2$ instead of $-m^2$, it will give rise to a wrong angular part in the total wave function, namely, $\exp(2im\varphi)$ instead of $\exp(im\varphi)$. Therefore, the separation procedure using parabolic coordinates is *not* invertible. It is three dimensional. Furthermore, since $\beta+\gamma=\varphi$, the phase angles are coupled and the hydrogen atom corresponds to a pair of coupled two-dimensional oscillators.

4. MATRIX ELEMENTS

The correspondence established in previous sections allows the 15 generators of SO(4,2) to admit the following realization in the oscillator representation:[15]

$$L_{23} = -\tfrac{1}{2}(A_-B^*_+ + A^*_-B_+ - A^*_+B_- - A_+B^*_-),$$
$$L_{31} = \tfrac{1}{2}i(A_-B^*_+ - A^*_-B_+ - A^*_+B_- + A_+B^*_-),$$
$$L_{12} = \tfrac{1}{2}(A^*_+A_+ - A^*_-A_- + B^*_+B_+ - B^*_-B_-),$$
$$L_{14} = \tfrac{1}{2}(A^*_-B_+ + A_-B^*_+ + A^*_+B_- + A_+B^*_-),$$
$$L_{24} = \tfrac{1}{2}i(A^*_-B_+ - A_-B^*_+ - A^*_+B_- + A_+B^*_-),$$
$$L_{34} = \tfrac{1}{2}(A^*_+A_+ + A^*_-A_- - B^*_+B_+ - B^*_-B_-),$$
$$L_{15} = \tfrac{1}{2}(A^*_+B^*_+ + A_-B_- + A_+B_+ + A^*_-B^*_-),$$
$$L_{25} = -\tfrac{1}{2}i(A^*_+B^*_+ + A_-B_- - A_+B_+ - A^*_-B^*_-),$$
$$L_{35} = \tfrac{1}{2}(A^*_+A^*_- + A_+A_- - B^*_+B^*_- - B_+B_-),$$
$$L_{45} = -\tfrac{1}{2}i(A^*_+A^*_- - A_+A_- + B^*_+B^*_- - B_+B_-),$$
$$L_{16} = \tfrac{1}{2}i(A^*_+B^*_- - A_+B_+ - A_-B_- + A^*_-B^*_+),$$
$$L_{26} = \tfrac{1}{2}(A^*_+B^*_+ + A_+B_+ - A_-B_- - A^*_-B^*_-),$$
$$L_{36} = \tfrac{1}{2}i(A^*_+A^*_- - A_+A_- - B^*_+B^*_- + B_+B_-),$$
$$L_{46} = \tfrac{1}{2}(A^*_+A^*_- + A_+A_- + B^*_+B^*_- + B_+B_-),$$
$$L_{56} = \tfrac{1}{2}(A^*_+A_+ + A^*_-A_- + B^*_+B_+ + B^*_-B_- + 2),$$

where

$$A_\pm = (a_1\mp ia_2)/2^{1/2}, \quad A^*_\pm = (a_1^+\pm ia_2^+)/2^{1/2},$$
$$B_\pm = (a_3\mp ia_4)/2^{1/2}, \quad B^*_\pm = (a_3^+\pm ia_4^+)/2^{1/2}$$

are the shift operators and

$$a_i = (\partial/\partial y_i + y_i)/2^{1/2}, \quad a_i^+ = (-\partial/\partial y_i + y_i)/2^{1/2}$$

are the boson operators satisfying the commutation relations

$$[a_i, a_j^+] = \delta_{ij},$$

with all other commutators vanishing. The generators T_1, T_2, T_3 given in Sec. 2 are, in SO(4,2) notation, L_{46}, L_{45}, L_{56}, respectively.

To calculate the mean value of r^p, we note that

$$r^p = s^{2p} = (n/2)^p y^{2p} = n^p(T_3-T_1)^p. \quad (36)$$

Then, in view of Eq. (24) and the normalization, we can write

$$\langle nl|r^p|nl\rangle = n^{p-1}\langle Nk|(T_3-T_1)^{p+1}|Nk\rangle. \quad (37)$$

Using Eq. (37), we have calculated the mean values for $p=-1,1,2,3,4,5$. The calculation is facilitated by the fact that only diagonal terms in $(T_3-T_1)^{p+1}$ contribute. With the use of commutation relations given in Eq. (18) and below

$$[T_+,T_-] = -2T_3, \quad [T_+,T_-] = \pm T_\pm, \quad (38)$$

and by denoting $P = T_+T_- + T_3$ and $Q = T_+T_-$, we have

$$(T_3 - T_1)^2 = T_3^2 + \tfrac{1}{2}P + \text{nondiagonal terms (nd.)},$$
$$(T_3 - T_1)^3 = T_3^3 + \tfrac{1}{2}T + (\tfrac{3}{2})PT_3 + (\text{nd.}),$$
$$(T_3 - T_1)^4 = T_3^4 + 2T^2 + (\tfrac{3}{4})P + (\tfrac{3}{4})PT_3 + (\tfrac{3}{8})Q^2 + (\text{nd.}),$$
$$(T_3 - T_1)^5 = T_3^5 + 5T_3^3 + (\tfrac{3}{2})T_3 + (\tfrac{25}{4})PT_3 + (\tfrac{15}{4})PT_3^2 + 5PT_3^3 + (\tfrac{15}{8})Q^2T_3 + \text{nd.},$$
$$(T_3 - T_1)^6 = T_3^6 + 10T_3^4 + (\tfrac{97}{8})T_3^2 + (\tfrac{15}{4})P + 5PT_3 + (\tfrac{225}{8})PT_3^2 + (\tfrac{45}{4})PT_3^3 + (\tfrac{15}{2})PT_3^4 + (\tfrac{5}{2})Q^2 + (\tfrac{15}{16})PQ^2 + (\tfrac{5}{8})QT_3^2 + (\tfrac{45}{4})Q^2T_3^2 + \text{nd.}$$

Using the above in Eq. (37) in conjunction with Eqs. (19) and (20), we have reproduced the explicit results obtained by the method of direct integration[16] up to $p = 4$. For $p = 5$, we obtain

$$\overline{r^5} = (\tfrac{231}{16})n^6 + (\tfrac{1065}{16})n^4 + (\tfrac{177}{8})n^2 - (\tfrac{305}{16})n^4l(l+1) - (\tfrac{525}{16})n^2l(l+1) + (\tfrac{35}{2})n^2l^2(l+1)^2 - (\tfrac{3}{2})l^3(l+1)^3 + (\tfrac{5}{2})l^2(l+1)^2 - (\tfrac{15}{4})l(l+1).$$

In comparing the results for different powers of p, we note a pattern common to all, namely, only even powers of n contribute.

We now deal with matrix elements with $n \neq n'$. As a specific example, we take the transition matrix element $\langle n'l'm'|z|nlm\rangle$. All other matrix elements can be similarly dealt with. To evaluate this matrix element in the oscillator representation, we utilize Eq. (32) by first making the expansion

$$\left|nlm\right\rangle = \sum_{n_1n_2} A_{nlm}^{n_1n_2} \left|n_1n_2m\right\rangle, \tag{39}$$

where $A_{nlm}^{n_1n_2}$ are the coefficients derived by Tarter.[17] Thus we have

$$\langle n'l'm'|z|nlm\rangle = \sum_{n_1n_2}\sum_{n_1'n_2'} A_{nlm}^{n_1n_2}A_{n'l'm'}^{n_1'n_2'} \times\left\langle n_1'n_2'm'\left|\frac{\xi^2-\eta^2}{8}\right|n_1n_2m\right\rangle. \tag{40}$$

It should be noted that the quantum numbers are related by $n = n_1 + n_2 + m + 1$. The left-hand side of Eq. (40) can be reduced to a radial integral in accordance with the selection rules $\Delta m = 0$, $\Delta l = \pm 1$. Denote the matrix element on the right-hand side by M. Then we have from Eq. (32)

$$M = \frac{2}{nn'}\left\langle N_1'N_2'm;\omega\left|S^*\left(\frac{2}{n'\omega}\right)\frac{u^4-v^4}{8}S\left(\frac{2}{n\omega}\right)\right| \times N_1N_2m;\omega\right\rangle. \tag{41}$$

Noting that in terms of the SO(4,2) generators $\frac{1}{4}(u^2+v^2) = (L_{46}+L_{56})/\omega$, $\frac{1}{4}(u^2-v^2) = (L_{34}+L_{35})/\omega$, $S^*(\lambda) = S(1/\lambda)$, and $S(1/\lambda)(L_{46}+L_{56})(L_{34}+L_{35}) = \lambda^{-2} \times (L_{46}+L_{56})(L_{34}+L_{35})S(1/\lambda)$, we can write

$$M = \frac{n'}{n}\left\langle N_1'N_2'm\left|(L_{46}+L_{56})(L_{34}+L_{35})S\left(\frac{n'}{n}\right)\right|N_1N_2m\right\rangle = \frac{n'}{n}\sum_{N_1''N_1''}\left\langle N_1'N_2'm\left|(L_{46}+L_{56})(L_{34}+L_{35})\right|N_1''N_2''m\right\rangle \times\left\langle N_1''N_2''m\left|S\left(\frac{n'}{n}\right)\right|N_1N_2m\right\rangle, \tag{42}$$

where the parameter ω has been omitted. This result has been obtained by Englefield[18] in parabolic coordinates making use of assumptions based upon physical insight. We have now provided a rigorous theoretical basis for its validity in the framework of a four-dimensional harmonic oscillator which has the same underlying symmetry as the hydrogen atom. The matrix elements in Eq. (42) can be evaluated algebraically. Putting Eq. (42) back into Eq. (40), we obtain the desired result, namely, an expression for the radial integral on the left-hand side in terms of matrix elements in the oscillator representation.

APPENDIX

The mapping from a three-dimensional space with Cartesian coordinates x_1, x_2, x_3 to a four-dimensional space with coordinates s_1, s_2, s_3, s_4 is bijective. It is tantamount to making a Kustaanheimo–Stiefel transformation[5]

$$\begin{pmatrix}x_1\\x_2\\x_3\\0\end{pmatrix} = \begin{pmatrix}s_3 & -s_4 & s_1 & -s_2\\ s_4 & s_3 & s_2 & s_1\\ s_1 & s_2 & -s_3 & -s_4\\ s_2 & -s_1 & -s_4 & s_3\end{pmatrix}\begin{pmatrix}s_1\\s_2\\s_3\\s_4\end{pmatrix} = A\begin{pmatrix}s_1\\s_2\\s_3\\s_4\end{pmatrix}, \tag{A1}$$

where A is the transformation matrix. If we designate the momenta conjugate to s_j as q_j, then we have

$$\begin{pmatrix}p_1\\p_2\\p_3\\0\end{pmatrix} = \frac{2}{r}A\begin{pmatrix}q_1\\q_2\\q_3\\q_4\end{pmatrix}. \tag{A2}$$

The transformation is contingent upon the relationships $s^2 = r$, $2\alpha = \theta$, and $\beta + \gamma = \varphi$, where r,θ,φ are spherical coordinates. For this choice of phase relationship, we require $m_1 = m$, $m_2 = 0$. Using the following

$$\frac{\partial}{\partial s_1} = \cos\beta\left(\cos\alpha\frac{\partial}{\partial s} - \frac{\sin\alpha}{s}\frac{\partial}{\partial\alpha}\right) - \frac{\sin\beta}{s\cos\alpha}\frac{\partial}{\partial\beta}, \tag{A3}$$
$$\frac{\partial}{\partial s_2} = \sin\beta\left(\cos\alpha\frac{\partial}{\partial s} - \frac{\sin\alpha}{s}\frac{\partial}{\partial\alpha}\right) + \frac{\cos\beta}{s\cos\alpha}\frac{\partial}{\partial\beta}, \tag{A4}$$
$$\frac{\partial}{\partial s_3} = \cos\gamma\left(\sin\alpha\frac{\partial}{\partial s} + \frac{\cos\alpha}{s}\frac{\partial}{\partial\alpha}\right) - \frac{\sin\gamma}{s\sin\alpha}\frac{\partial}{\partial\gamma}, \tag{A5}$$
$$\frac{\partial}{\partial s_4} = \sin\gamma\left(\sin\alpha\frac{\partial}{\partial s} + \frac{\cos\alpha}{s}\frac{\partial}{\partial\alpha}\right) + \frac{\cos\gamma}{s\sin\alpha}\frac{\partial}{\partial\gamma}, \tag{A6}$$

we can show that the last component of Eq. (A2),

$$s_2q_1 - s_1q_2 - s_4q_3 + s_3q_4 = -i(\partial/\partial\beta - \partial/\partial\gamma) = 2N_z = 0, \tag{A7}$$

is consistent with our choice of m_1 and m_2. On the other hand, using Eqs. (A3)–(A6), we can show that $K_z = L_{12}$, $K_\pm = L_{23} \pm iL_{31}$. If we choose the other phase relationship, namely $\beta - \gamma = \varphi$, then the sign of s_4 in the transformation matrix A must be reversed and the roles of N_z and K_z are interchanged.[19] In other words, N_z and $N_\pm$ will then become the angular momentum operators instead. Since $\gamma \to -\gamma$ means $s_4 \to -s_4$, by changing the sign of s_4 (or y_4) in the generators, we obtain an alternative representation of the dynamical group.

Equations (A3)–(A6) can be used to show the equivalence between the generators in the oscillator representation and those obtained by Barut and Bornzin.[20] For example,

$$\begin{aligned} W_1 &= L_{16} = -\tfrac{1}{2}i(-s_1\partial/\partial s_3 - s_3\partial/\partial s_1 \\ &\quad + s_2\partial/\partial s_4 + s_4\partial/\partial s_2) \\ &= i(-\sin\theta\cos\varphi r\partial/\partial r - \cos\varphi\cos\theta\partial/\partial\theta \\ &\quad + \sin\varphi/\sin\theta\partial/\partial\varphi) \\ &= -ir\partial/\partial x_1 = rp_1. \end{aligned}$$

This shows that the extra factors $e^{\pm i\varphi}$ are not needed as in Appendix E of Ref. 7.

[1]E. Schrödinger, Proc. R. Irish Acad. Sect. A **46**, 183 (1941).
[2]E. Chacón, D. Levi, and M. Moshinsky, J. Math. Phys. **17**, 1919 (1976).
[3]C. Quesne and M. Moshinsky, J. Math. Phys. **12**, 1780 (1971).
[4]J. Čížek and J. Paldus, Int. J. Quantum Chem. **12**, 875 (1977).
[5]A. O. Barut, C. K. E. Schneider, and R. Wilson, J. Math. Phys. **20**, 2244 (1979).
[6]A. C. Chen, Phys. Rev. A **22**, 333, 2901E (1980).
[7]M. J. Englefield, *Group Theory and the Coulomb Problem* (Wiley-Interscience, New York, 1972).
[8]In treating the Stark effect by perturbation theory in parabolic coordinates, the separated perturbed equations are compared with the equation of a two-dimensional anharmonic oscillator and the separated unperturbed equations with that of a two-dimensional oscillator.
[9]S. Graffi and V. Grecchi, Commun. Math. Phys. **62**, 83 (1978). These authors pointed out that "subtleties arise in the inversion of this procedure." F. Ravndal and T. Toyoda, Nucl. Phys. B **3**, 312 (1967). These authors noted the danger in not properly treating the symmetry properties of the system and introduced phase angles φ_μ and φ_ν for the separated equations so that $\varphi = \varphi_\mu \pm \varphi_\nu$.
[10]See, for example, B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[11]See Ref. 10.
[12]V. Fock, Z. Phys. **98**, 145 (1935).
[13]Appendix C of Ref. 7.
[14]In squared parabolic coordinates $u^2 = \xi$, $v^2 = \eta$, the separated equations have the form $[-d^2/du^2 - Eu^2 + (m^2 - \frac{1}{4})u^{-2}]f(u^2) = 2Z_1 f(u^2)$. The angular dependence is also missing.
[15]Compare with Appendix E of Ref. 7.
[16]H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1967).
[17]C. B. Tarter, J. Math. Phys. **11**, 3192 (1970).
[18]Page 99 of Ref. 7.
[19]For $\beta - \gamma = \varphi$, we have $m_1 = 0$, $m_2 = m$, and $d^l_{0m}(\theta)\exp(im\varphi) = (-1)^m \times [4\pi/(2l+1)]^{1/2} Y_{lm}(\theta,\varphi)$.
[20]A. O. Barut and G. L. Bornzin, J. Math. Phys. **12**, 841 (1971).

Simple multiple explode-decay mode solutions of a two-dimensional nonlinear Schrödinger equation

Akira Nakamura
Department of Mathematics, Clarkson College of Technology, Potsdam, New York 13676

(Received 29 June 1981; accepted for publication 16 October 1981)

Multiple similarity type explode-decay mode solutions of a two-dimensional (≡2D) nonlinear Schrödinger (≡NLS) equation have been obtained by the bilinear method. From the three examples of the 2D–KdV equation, ordinary cubic 2D–NLS equation, and the present 2D–NLS equation, the expectation is presented such that "any 2D nonlinear evolution equation which has multiple soliton solutions simultaneously has simple self-similar-type multiple explode–decay mode solutions so far as the equation has self-similar symmetry."

PACS numbers: 03.65.Ge

1. INTRODUCTION

It is known that one of the typical nonlinear evolution equations, the 2-dimensional (≡2D) KdV equation, has not only well-known multiple soliton solutions but also self-similar-type simple multiple explode-decay mode solutions[1–5] which we call in this paper ripplons for brevity. The equation is written as

$$(u_t + 6uu_x + u_{xxx})_x + 3\alpha^2 u_{yy} = 0. \tag{1.1}$$

Throughout this paper we use the following notations. Subscripts x, y, t represent partial differentiation, $\alpha, \beta, \beta', \delta$ represent arbitrary real constants, and * represents complex conjugate. These ripplon solutions are interesting since they have precisely the same nonlinear superposition property as solitons.[4,5]

Recently, by generalizing the similarity type plane wave solution derived by Redekopp, a similar type (ripplon) solution has been found for a 2D cubic nonlinear Schrödinger (≡NLS) equation,[6]

$$iu_t + \beta u_{xx} + \beta' u_{yy} + \delta u^* uu = 0. \tag{1.2}$$

However, in this case just as multiple soliton solutions exist essentially only in the 1D direction, multiple ripplon solutions also exist essentially only in the 1D direction.

As the third example of multiple ripplon solutions, we report here the ones in another 2D–NLS equation,[7–9] written as

$$iu_t - \beta u_{xx} + \beta' u_{yy} + \delta u^* uu - 2wu = 0, \tag{1.3a}$$

$$\beta w_{xx} + \beta' w_{yy} - \beta\delta(u^* u)_{xx} = 0. \tag{1.3b}$$

This equation is known to fit into inverse scattering transform formalism,[10] thus to have multiple soliton solutions which are nonlinear superpositions of 1D-like solitons in 2D arbitrary propagating directions.[11,12] Then in contrast to the rather limited superposition properties of Eq. (1.2), we can expect the existence of the multiple ripplon solutions of similar 2D character to Eqs. (1.3a) and (1.3b). As will be shown in this paper, this turns out to be the case.

Having concrete examples of these three equations, we are naturally led to the following empirical proposition: *Any two (space) dimensional nonlinear evolution equation which has multiple soliton solutions has, simultaneously, the multiple simple self-similar-type explode-decay mode (ripplon) solutions so far as the equation has self-similar symmetry.*

Now returning to Eqs. (1.3a) and (1.3b), we present explicit multiple ripplon solutions in the following sections.

2. BILINEAR FORM AND 1-RIPPLON SOLUTIONS

For our analysis, we use the technique of the Hirota bilinear method.[13,14] We consider the dependent variable transform[12]

$$u = g/f, \quad w = (2\beta \ln f)_{xx}, \quad (f = f^*), \tag{2.1}$$

Inserting this into Eqs. (1.3a) and (1.3b) we have

$$(iD_t + (-\beta)D_x^2 + \beta' D_y^2)g\cdot f = 0, \tag{2.2a}$$

$$(\beta D_x^2 + \beta' D_y^2) f\cdot f - \delta g^* g = 0. \tag{2.2b}$$

Here bilinear differential operators are defined for arbitrary functions $a(x, y, t)$ and $b(x, y, t)$ as[13,14]

$$\begin{aligned} &D_x^{\,l} D_y^{\,m} D_t^{\,n} a(x,y,t)\cdot b(x,y,t) \\ &\quad \equiv (\partial_x - \partial_{x'})^l (\partial_y - \partial_{y'})^m (\partial_t - \partial_{t'})^n \\ &\qquad \times a(x,y,t) b(x',y',t')|_{x'=x, y'=y, t'=t}. \end{aligned} \tag{2.3}$$

Quite similarly to the recent result[6] of the cubic 2D–NLS Equation (1.2), we can directly check that the following 1-ripplon solution satisfies Eqs. (2.2a) and (2.2b):

$$g = (1/t)e^{\theta_i}, \qquad f = 1 + e^{\theta_i + \theta_i^* + \tau_{ii^*}}, \tag{2.4}$$

$$\theta_i \equiv \frac{i(x + x_i)^2}{4(-\beta)t} + \frac{i(y + y_i)^2}{4\beta' t} + v_i,$$

$$e^{-\tau_{ii^*}} \equiv \frac{[(ix_i) + (ix_i)^*]^2}{2\beta\delta} + \frac{[(iy_i) + (iy_i)^*]^2}{2\beta'\delta}, \tag{2.5}$$

where x_i, y_i, v_i, are arbitrary complex constants. In the original variable u the solution is written as

$$\begin{aligned} u = \frac{g}{f} &= \frac{\exp[-(1/2)\tau_{ii^*}]}{2t} \operatorname{sech} \frac{1}{2t}\left[\frac{2i(x_i - x_i^*)x + ix_i^2 - ix_i^{*2}}{4(-\beta)}\right. \\ &\quad \left. + \frac{2i(y_i - y_i^*)y + iy_i^2 - iy_i^{*2}}{4\beta'} + (v_i + v_i^* + \tau_{ii^*})t\right] \\ &\quad \times \exp\left[i\frac{(x + x_i)^2 + (x + x_i^*)^2}{8(-\beta)t}\right. \\ &\quad \left. + i\frac{(y + y_i)^2 + (y + y_i^*)^2}{8\beta' t} + \frac{v_i - v_i^*}{2}\right]. \end{aligned} \tag{2.6}$$

 0022-2488/82/030417-04$02.50 © 1982 American Institute of Physics

We note that except for the change of sign $\beta \leftrightarrow (-\beta)$, this form is quite similar to the 1-ripplon solution of the ordinary 2D cubic nonlinear Schrödinger equation (1.2), and has similar characteristics as follows: The solution is essentially one-dimensionally localized. It grows its amplitude and narrows its width in time $(-\infty < t < 0)$ and then decays its amplitude and broadens its width in time $(0 < t < \infty)$. At $t = 0$, the solution explodes to infinity as a 1D Dirac δ function. From this characteristic, the present ripplon solution may be considered as a kind of essentially 1D version of instanton solution. The propagating velocity is constant during the whole process.

3. N-RIPPLON SOLUTIONS

We note that in Eq. (2.4), apart from the additional factor $1/t$ in g, the 1-ripplon solution is formally quite similar to the 1-soliton solution written as

$$g = e^{\eta_i}, \quad f = 1 + \exp(\eta_i + \eta_i^* + \bar{\tau}_{ii^*}),$$
$$\eta_i \equiv k_i x + l_i y + \omega_i t, \quad i\omega_i + (-\beta)k_i^2 + \beta' l_i^2 = 0,$$
$$\exp(-\bar{\tau}_{ii^*}) \equiv \frac{2\beta}{\delta}(k_i + k_i^*)^2 + \frac{2\beta'}{\delta}(l_i + l_i^*)^2. \quad (3.1)$$

This formal similarity between 1-ripplon and 1-soliton solutions leads us to the natural expectation that N-ripplon solutions may exist and have the form analogous to the N-soliton solutions. Thus in direct analogy to the N-soliton case, we have the assumed form of N-ripplon solutions as

$$g = \sum_{n=0,1}^{(1)} \frac{1}{t} \exp\left\{\sum_{i=1}^{2N} n_i\theta_i + \sum_{i<j=1}^{2N} n_i n_j \tau_{ij}\right\}, \quad (3.2a)$$

$$f = \sum_{n=0,1}^{(0)} \exp\left\{\sum_{i=1}^{2N} n_i\theta_i + \sum_{i<j=1}^{2N} n_i n_j \tau_{ij}\right\}, \quad (3.2b)$$

where $\theta_i (1 \leqslant i \leqslant N)$ is the same as that given in Eq. (2.5) and $n_i = 0$ or $1 (1 \leqslant i \leqslant 2N)$, $\theta_{N+i} \equiv (\theta_i)^*$,

$$e^{\tau_{ij}} \equiv \frac{(ix_i - ix_j)^2}{2\beta\delta} + \frac{(iy_i - iy_j)^2}{2\beta'\delta}, \quad e^{\tau_{N+i,N+j}} \equiv (e^{\tau_{ij}})^*,$$
$$e^{-\tau_{i,N+j}} \equiv e^{-\tau_{N+j,i}} \equiv e^{-\tau_{ij^*}}$$
$$\equiv \frac{[(ix_i) + (ix_j)^*]^2}{2\beta\delta} + \frac{[(iy_i) + (iy_j)^*]^2}{2\beta'\delta}$$
$$(1 \leqslant i, j \leqslant N), \quad (3.3)$$

and the summation symbol $\Sigma_{n=0,1}^{(m)}$ implies that in the total 2^{2N} sums over $n_1 = 0,1,\ldots,n_{2N} = 0,1$, only those sums satisfying the condition

$$(n_1 + n_2 + \cdots + n_N) - (n_{N+1} + n_{N+2} + \cdots + n_{2N}) = m \quad (3.4)$$

should be taken. As the simplest such important example of multiple ripplons, we write explicit forms of f and g for the 2 ripplon $(N = 2)$ case

$$g = (1/t)[\exp(\theta_1) + \exp(\theta_2) + \exp(\theta_1 + \theta_2 + \theta_1^* + \tau_{121^*})$$
$$\exp(\theta_1 + \theta_2 + \theta_2^* + \tau_{122^*})],$$
$$f = 1 + \exp(\theta_1 + \theta_1^* + \tau_{11^*}) + \exp(\theta_1 + \theta_2^* + \tau_{12^*})$$
$$+ \exp(\theta_2 + \theta_1^* + \tau_{21^*}) + \exp(\theta_2 + \theta_2^* + \tau_{22^*})$$
$$+ \exp(\theta_1 + \theta_2 + \theta_1^* + \theta_2^* + \tau_{121^*2^*}),$$
$$\tau_{121^*} \equiv \tau_{12} + \tau_{11^*} + \tau_{21^*}, \quad \tau_{122^*} \equiv \tau_{12} + \tau_{12^*} + \tau_{22^*}$$
$$\tau_{121^*2^*} \equiv \tau_{12} + \tau_{11^*} + \tau_{12^*} + \tau_{21^*} + \tau_{22^*} + \tau_{1^*2^*}.$$

The symmetrized form of this g and f may be obtained in a manner similar to the case of the 1D–NLS equation given in Appendix B of Ref. 13. From here to the end of this section we give the proof that Eqs. (3.2a) and (3.2b) satisfy bilinear equations (2.2a) and (2.2b) and therefore actually are the solutions of the 2D–NLS Equation (1.3). The readers who are interested only in basic physical aspects can go directly to the next section.

For the purpose of the proof, we recall that in the case of the usual solitons the bilinear method utilizes the simple relation between the arbitrary bilinear operator $F(D_x, D_y, D_t)$ and simple exponential functions

$$F(D_x, D_y, D_t)\exp(kx + ly + \omega t)\cdot\exp(k'x + l'y + \omega' t)$$
$$= F(k - k', l - l', \omega - \omega')\exp(kx + ly + \omega t)$$
$$\times \exp(k'x + l'y + \omega' t). \quad (3.5)$$

In other words, simple exponential functions are eigenfunctions of the bilinear differential operators. In the present case, we must seek the corresponding relation for $\exp(\theta_i)$. By the direct calculation we obtain the following relations:

$$(iD_t + (-\beta)D_x^2 + \beta' D_y^2)\frac{1}{t}\exp\left(\sum_{i=1}^{2N} n_i\theta_i\right)\cdot\exp\left(\sum_{i=1}^{2N} n_i'\theta_i\right)$$
$$= \left\{\frac{-1}{4(-\beta)t^2}\left[\sum_{i=1}^{N}(n_i - n_i')(ix_i)^2\right.\right.$$
$$- \sum_{i=1}^{N}(n_{N+i} - n_{N+i}')(ix_i)^{*2}$$
$$\left.- \left(\sum_{i=1}^{N}(n_i - n_i')(ix_i) + \sum_{i=1}^{N}(n_{N+i} - n_{N+i}')(ix_i)^*\right)^2\right]$$
$$+ [\text{terms with interchange of}$$
$$(-\beta) \to \beta', x \to y, x_i \to y_i, x_i^* \to y_i^*]\Bigg\}$$
$$\times \frac{1}{t}\exp\left[\sum_{i=1}^{2N}(n_i + n_i')\theta_i\right], \quad (3.6)$$

$$(\beta D_x^2 + \beta' D_y^2)\exp\left(\sum_{i=1}^{2N} n_i'\theta_i\right)\cdot\exp\left(\sum_{i=1}^{2n} n_i''\theta_i\right)$$
$$= \left\{\frac{1}{4\beta t^2}\left[\sum_{i=1}^{N}(n_i' - n_i'')(ix_i)\right.\right.$$
$$\left.+ \sum_{i=1}^{N}(n_{N+i}' - n_{N+i}'')(ix_i)^*\right]^2$$
$$+ \frac{1}{4\beta' t^2}\left[\sum_{i=1}^{N}(n_i' - n_i'')(iy_i)\right.$$
$$\left.\left.+ \sum_{i=1}^{N}(n_{N+i}' - n_{N+i}'')(iy_i)^*\right]^2\right\}$$
$$\times \exp\left[\sum_{i=1}^{2N}(n_i' + n_i'')\theta_i\right], \quad (3.7)$$

where $n_i, n_i', n_i'' (1 \leqslant i \leqslant 2N)$ take the values 0 or 1 and satisfy the conditions,

$$\sum_{i=1}^{N}(n_i - n_{N+i}) = 1, \quad \sum_{i=1}^{N}(n_i' - n_{N+i}') = 0,$$
$$\sum_{i=1}^{N}(n_i'' - n_{N+i}'') = 0. \quad (3.8)$$

Equations (3.6) and (3.7) show that the functions $\exp(\theta_i)$ again have the nice property of being eigenfunctions of the bilinear differential operators. We note that the one small complication is that the eigenvalues now contain an addi-

tional time-dependent factor t^{-2} everywhere.

Now we insert the assumed N-ripplon solution, Eqs. (3.2a) and (3.2b), into the bilinear equations (2.2a) and (2.2b). From the above eigenfunction properties, after operations of bilinear differentiations, the results should again be the sum of products of $\exp(\theta_i)$'s with proper change of their coefficients. Now we will see those coefficients in detail. In the resultant two equations, let the coefficient of

$$\exp\{(\theta_1+\cdots+\theta_l)+2(\theta_{l+1}+\cdots+\theta_{l+m}) + (\theta_{N+1}+\cdots+\theta_{N+l'}) + 2(\theta_{N+l'+1}+\cdots+\theta_{N+l'+m'})\}$$

be denoted as $D_1(l,l')$ and $D_2(l,l')$, respectively. Using the relations (3.6) and (3.7), we see that

$$D_1(l,l')\propto \sum_{\sigma=0,1}{}^{(+1)}\left\{\frac{1}{4(-\beta)t^2}\left[\sum_{i=1}^{l}\sigma_i(ix_i)^2-\sum_{i=1}^{l'}\sigma_{N+i}(ix_i)^{*2}-\left(\sum_{i=1}^{l}\sigma_i(ix_i)+\sum_{i=1}^{l'}\sigma_{N+i}(ix_i)^*\right)^2\right]\right.$$
$$\left.+\ [\text{same term with } (-\beta) \text{ replaced by } \beta', x_i \text{ by } y_i, \text{ and } x_i^* \text{ by } y_i^*]\right\}$$
$$\times\frac{1}{t}\exp\Bigg(\sum_{\substack{i<j\\ 1<i,j<l\\ N<i,j<N+l'}}\frac{\sigma_i\sigma_j}{2}\tau_{ij}\Bigg), \tag{3.9}$$

$$D_2(l,l')\propto \sum_{\sigma=0,1}{}^{(0)}\left\{\frac{1}{4\beta t^2}\left[\sum_{i=1}^{l}\sigma_i(ix_i)+\sum_{i=1}^{l'}\sigma_{N+i}(ix_i)^*\right]^2+\frac{1}{4\beta' t^2}\left[\sum_{i=1}^{l}\sigma_i(iy_i)+\sum_{i=1}^{l'}\sigma_{N+i}(iy_i)^*\right]^2\right\}$$
$$\times\exp\Bigg(\sum_{\substack{i<j\\ 1<i,j<l\\ N<i,j<N+l'}}\frac{\sigma_i\sigma_j}{2}\tau_{ij}\Bigg)-\sum_{\sigma=0,1}{}^{(2)}\frac{\delta}{t^2}\exp\Bigg(\sum_{\substack{i<j\\ 1<i,j<l\\ N<i,j<N+l'}}\frac{\sigma_i\sigma_j}{2}\tau_{ij}\Bigg), \tag{3.10}$$

where $\sigma_i=+1$ or -1 $(i=1,\ldots,l,N+1,\ldots,N+l')$ and symbol $\Sigma_{\sigma=0,1}{}^{(m)}$ implies that in the total $2^{l+l'}$ sums over $\sigma_1=\pm1$; $\sigma_2=\pm1;\cdots$, only those sums satisfying the condition

$$\sum_{i=1}^{l}\sigma_i-\sum_{i=1}^{l'}\sigma_{N+i}=m \tag{3.11}$$

should be taken.

In Eqs. (3.9) and (3.10), inserting the expression for $\exp(\tau_{ij})$ given by Eq. (3.3) and then rewriting the notations by replacing

$$\sigma_i \text{ by } \hat\sigma_i,\quad -\sigma_{N+i} \text{ by } \hat\sigma_{l+i},$$
$$\frac{ix_i}{(2\beta\delta)^{1/2}} \text{ by } \hat x_i,\quad \frac{(ix_i)^*}{(2\beta\delta)^{1/2}} \text{ by } -\hat x_{l+i},$$
$$\frac{iy_i}{(2\beta'\delta)^{1/2}} \text{ by } \hat y_i,\quad \frac{(iy_i)^*}{(2\beta'\delta)^{1/2}} \text{ by } -\hat y_{l+i}, \tag{3.12}$$

we have

$$D_1(l,l')\propto \sum_{\hat\sigma=0,1}{}^{(1)}\Big\{-\sum_{i=1}^{n}\hat\sigma_i\hat x_i{}^2+\Big(\sum_{i=1}^{n}\hat\sigma_i\hat x_i\Big)^2+\sum_{i=1}^{n}\hat\sigma_i\hat y_i{}^2 - \Big(\sum_{i=1}^{n}\hat\sigma_i\hat y_i\Big)^2\Big\}b(\hat x,\hat y,\hat\sigma)\equiv D_1(n),$$

$$D_2(l,l')\propto \sum_{\hat\sigma=0,1}{}^{(0)}\Big\{\Big(\sum_{i=1}^{n}\hat\sigma_i\hat x_i\Big)^2+\Big(\sum_{i=1}^{n}\hat\sigma_i\hat y_i\Big)^2\Big\}b(\hat x,\hat y,\hat\sigma) - \sum_{\hat\sigma=0,1}{}^{(2)}2b(\hat x,\hat y,\hat\sigma)\equiv D_2(n),$$

$$b(\hat x,\hat y,\hat\sigma)\equiv\prod_{i<j=1}^{n}\{(\hat x_i-\hat x_j)^2+(\hat y_i-\hat y_j)^2\}^{(1+\hat\sigma_i\hat\sigma_j)/2}, \tag{3.13}$$

where $l+l'=n$ and $\Sigma_{\sigma=0,1}{}^{(m)}$ implies that in the total 2^n sums over $\hat\sigma_1=\pm1,\ldots,\hat\sigma_n=\pm1$, only those sums satisfying the condition

$$\sum_{i=1}^{n}\hat\sigma_i=m \tag{3.14}$$

should be taken.

Thus, expressions (2.7a) and (2.7b) are exact solutions provided that the following identities hold:

$$D_1(n)=0,\quad \text{for odd } n$$
$$D_2(n)=0,\quad \text{for even } n. \tag{3.15}$$

The quantities $D_i(n)(i=1,2)$ have the following properties:

(i) $D_i(n)$ is the polynomial of $\hat x_1\cdots\hat x_n\ \hat y_1\cdots\hat y_n$,

(ii) $D_i(n)$ is the invariant under the interchange of $(\hat x_i,\hat y_i)\leftrightarrow(\hat x_j,\hat y_j)$,

(iii) $D_i(n|\hat x_1=\hat x_2,\hat y_1=\hat y_2)\propto D_i(n-2|\hat x_3\cdots\hat x_n,\hat y_3\cdots\hat y_n)$.

The identities $D_1(1)=0$ and $D_2(0)=D_2(2)=0$ can be checked directly. Now we assume that the identities hold for $n-2$. Then from (ii) and (iii), $D_i(n)$ must contain factors $\{(\hat x_1-\hat x_2)^2$ or $(\hat y_1-\hat y_2)^2$ or $(\hat x_1-\hat x_2)(\hat y_1-\hat y_2)\}$ as shown in the Appendix. From (ii), contribution of these factors from each pair i,j $(1<i,j<N)$ leads to the inequality

$$[\text{order of } D_i(n)]>2_nC_2=n^2-n. \tag{3.16}$$

On the other hand, Eqs. (3.13) show that the

$$[\text{order of } D_1(n)]<2_{n+1/2}C_2+2_{n-1/2}C_2+2 = \frac{(n-1)^2}{2}+2,$$

$$[\text{order of } D_2(n)]<\text{larger of } (2_{n/2}C_2+2_{n/2}C_2+2) \text{ and } (2+{}_{n/2+1}C_2+2_{n/2-1}C_2) = \tfrac{1}{2}n^2-n+2. \tag{3.17}$$

From Eqs. (3.16) and (3.17) we see that $D_1(n)=0$ for $n>3$ and $D_2(n)=0$ for $n>4$ and the identities (3.15) are proved. Thus we have proved by the standard application of the bilinear

method that the 2D–NLS Equation (1.3) has simple multiple ripplon solutions given by Eqs. (2.1), (3.2a), and (3.2b).

4. PROPERTIES OF MULTIPLE RIPPLON SOLUTIONS

Since formally the N-ripplon solution is quite similar to the usual N-soliton solution, one can easily expect that quite parallel arguments to the usual N-soliton case hold everywhere. One physically interesting point is whether the phase shift occurring in the collision of two ripplons is time-dependent or not. Since all the factors $\exp(\tau_{ij})$ which are related to the phase shift are also time-independent in the present ripplon case, we see that the ripplon phase shift is time-independent.

ACKNOWLEDGMENTS

The present work was partially supported by the Mathematics Division of AFOSR and ONR.

APPENDIX

We have the following properties (I)–(III) for arbitrary finite-order polynomials.

(I) Any finite-order polynomial $P(x_1, x_2)$ can be decomposed as

$$P(x_1, x_2) = (x_1 + x_2)A(x_1, x_2) + (x_1 - x_2)B(x_1, x_2) + c_0, \tag{A1}$$

where $A(x_1, x_2)$, $B(x_1, x_2)$ are again finite polynomials and c_0 is constant with respect to x_1 and x_2.

Proof: Clearly we can decompose $P(x_1, x_2)$ as $P(x_1, x_2) = x_1\overline{A}(x_1, x_2) + x_2\overline{B}(x_1, x_2) + c_0$, where $\overline{A}(x_1, x_2)$, $\overline{B}(x_1, x_2)$ are finite polynomials. Then we have $P(x_1, x_2) = \frac{1}{2}(x_1 + x_2)(\overline{A} + \overline{B}) + \frac{1}{2}(x_1 - x_2)(\overline{A} - \overline{B}) + c_0$, which is equivalent to (A1).

(II) Denote an arbitrary finite polynomial of order at most n as P_n. $P_n(x_1, x_2, y_1, y_2)$ can be written as the sum

$$P_n(x_1, x_2, y_1, y_2) = \sum_{\substack{i,j,k,l \\ i+j+k+l \le n}} c_{ijkl}(x_1 + x_2)^i(x_1 - x_2)^j(y_1 + y_2)^k(y_1 - y_2)^l, \tag{A2}$$

where c_{ijkl} are constants with respect to x_1, x_2, y_1, y_2.

Proof: We use mathematical induction. For $n = 1$, we see (A2) holds as

$$c_1x_1 + c_2x_2 + c_3y_1 + c_4y_2 + c_0 = \tfrac{1}{2}(c_1 + c_2)(x_1 + x_2) + \tfrac{1}{2}(c_1 - c_2)(x_1 - x_2) + \tfrac{1}{2}(c_3 + c_4)(y_1 + y_2) + \tfrac{1}{2}(c_3 - c_4)(y_1 - y_2) + c_0,$$

where $c_0, \ldots, c_4$ are all constants. Next assume that it holds for P_n. By applying (I) twice, we have

$$\begin{aligned} P_{n+1}(x_1, x_2, y_1, y_2) &= (x_1 + x_2)A(x_1, x_2, y_1, y_2) + (x_1 - x_2)B(x_1, x_2, y_1, y_2) + C(y_1, y_2) \\ &= (x_1 + x_2)A(x_1, x_2, y_1, y_2) + (x_1 - x_2)B(x_1, x_2, y_1, y_2) \\ &\quad + (y_1 + y_2)D(y_1, y_2) + (y_1 - y_2)E(y_1, y_2) + c_0, \end{aligned}$$

where by construction A, B, D, E are polynomials of order at most n and c_0 is constant with respect to x_1, x_2, y_1, y_2. From our assumptions, A, B, D, E are sums of (A2) type; therefore $P_{n+1}(x_1, x_2, y_1, y_2)$ is again a sum of (A2) type with n replaced by $n + 1$. Therefore (A2) holds for all positive integers n.

(III) Suppose the polynomial $P_n(x_1, x_2, y_1, y_2)$ satisfies the conditions

$$P_n(x_1, x_2, y_1, y_2)|_{x_1 = x_2,\, y_1 = y_2} \equiv 0, \tag{A3}$$

$$P_n(x_1, x_2, y_1, y_2) \equiv P_n(x_2, x_1, y_2, y_1), \tag{A4}$$

where the symbol $\equiv$ represents "identically equal to." Then

$$P_n(x_1, x_2, y_1, y_2) = (x_1 - x_2)^2A + (x_1 - x_2)(y_1 - y_2)B + (y_1 - y_2)^2C, \tag{A5}$$

where A, B, C are appropriate finite polynomials.

Proof: We consider the expansion (A2) for $P_n(x_1, x_2, y_1, y_2)$ proved in (II). Then from condition (A3), $c_{ijkl} = 0$ for $j = 0$, $l = 0$. From conditions (A4), $c_{ijkl} = 0$ for $j + l =$ odd. Therefore in the expansion (A2), the nonvanishing terms are at most $j + l = 2, 4, 6, \cdots$. Namely,

$$\begin{aligned} P_n(x_1, x_2, y_1, y_2) &= a_1(x_1 - x_2)^2 + a_2(x_1 - x_2)^1(y_1 - y_2)^1 + a_3(y_1 - y_2)^2 \\ &\quad + b_1(x_1 - x_2)^4 + b_2(x_1 - x_2)^3(y_1 - y_2)^1 \\ &\quad + b_3(y_1 - y_2)^2(y_1 - y_2)^2 + \cdots, \end{aligned}$$

where $a_1, a_2, \ldots b_1, b_2, \ldots$, are constants, Thus we have (A5).

[1] R. S. Johnson and S. Thompson, Phys. Lett. A **66**, 279 (1978).
[2] N. C. Freeman, Adv. Appl. Mech. **20**, 1 (1980).
[3] L. G. Reddekopp. Stud. Appl. Math. **63**, 185 (1980).
[4] A. Nakamura, Phys. Rev. Lett. **46**, 751 (1981).
[5] A. Nakamura, J. Math. Phys. **22**, 2456 (1981).
[6] A. Nakamura, J. Phys. Soc. Jpn. **50**, 2467 (1981).
[7] D. J. Benney and G. J. Roskes, Stud. Appl. Math. **48**, 377 (1969).
[8] A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).
[9] N. C. Freeman and A. Davey, Proc. R. Soc. London Ser. A **344**, 427 (1975).
[10] M. J. Ablowitz and R. Haberman, Phys. Rev. Lett. **35**, 1185 (1975).
[11] D. Anker and N. C. Freeman, Proc. R. Soc. London Ser. A **360**, 529 (1978).
[12] J. Satsuma and M. J. Ablowitz, J. Math. Phys. **20**, 1496 (1979).
[13] R. Hirota, J. Math. Phys. **14**, 805 (1973).
[14] R. Hirota and J. Satsuma, Prog. Theor. Phys. Suppl. **59**, 64 (1976).

A combinatoric result related to the N-body problem [a)]

Bengt R. Karlsson
Tandem Accelerator Laboratory, Box 533, S-75121 Uppsala, Sweden

(Received 18 August 1981; accepted for publication 2 October 1981)

In the set of complete chains of partitions of N objects, let chains that are related through a permutation of the objects be termed equivalent. The number of equivalence classes μ_N is shown to equal the Euler number $|E_{N-1}|$ if N is odd, and $2^N(2^N-1)|B_N|/N$, where B_N is a Bernoulli number, if N is even. The number of elements in each class is also found. In the Yakubovskii-type formulation of the N-body problem in quantum mechanics, μ_N is the basic number of coupled equations when all particles are identical.

PACS numbers: 03.65.Nk, 02.10. + w

In current formulations of the quantum mechanical N-body (scattering) problem, the concept of a partition of N particles into k groups b_k is frequently used. A sequence of succesive partitions such that $b_2 \supset b_3 \supset \cdots \supset b_{N-1} \supset b_N$ is called a complete (or maximal) chain. In the most detailed formulation, that of Yakubovskii,[1] such chains appear as the index set for the quantities of interest, like transition operators or wavefunction components.[2]

It is known that the total number of chains is $N!(N-1)!/2^{N-1}$, and this is also the number of coupled equations in the Yakubovskii-type formulation, for distinguishable particles.[3] For identical particles, chains that can be transformed into each other through a permutation of the particles are equivalent, and the number of coupled equations reduces to the number μ_N of equivalence classes.[4] As is easily verified, $\mu_3 = 1$ and $\mu_4 = 2$. A general expression for μ_N has not been obtained (see Ref. 3), and we therefore proceed to derive such an expression.

In a k-particle chain, $k-1$ splittings are involved. Therefore, in order to produce an N-particle chain starting with the partition $(1,...,k)(k+1,...,N)$ an additional $N-2$ splittings are required, $k-1$ of which are splittings of the set $(1,...,k)$. These last splittings can be carried out anywhere along the N-particle chain, and therefore in $\binom{N-2}{k-1}$ different ways. Each choice corresponds to a different class of N-particle chains, and the total number of classes is therefore $(N \geqslant 3; \mu_1 = \mu_2 = 1)$

$$\mu_N = \frac{1}{2}\sum_{k=1}^{N-1}\binom{N-2}{k-1}\mu_k\mu_{N-k}. \tag{1}$$

The factor $\frac{1}{2}$ takes care of the fact that $(1,...,k)(k+1,...,N)$ and $(1,...,N-k)(N-k+1,...,N)$ generate the same classes. In order to relate μ_N to familiar combinatoric numbers, consider

$$g(s) \equiv \sum_{N=1}\mu_N s^{N-1}/(N-1)!. \tag{2}$$

Using (1) (for $N \geqslant 3$), and interchanging the order of summation, we find

$$g(s) = \mu_1 + \mu_2 - \tfrac{1}{2}s\mu_1^2 + \frac{1}{2}\sum_{k=1}^{\infty}\sum_{N=k+1}^{\infty}\mu_k\mu_{N-k}s^{N-1}/((N-1)\times(k-1)!(N-k-1)!). \tag{3}$$

Replacing N by $i = N-k$, and using $\mu_1 = \mu_2 = 1$, it is seen that

$$\frac{d}{ds}g(s) = \tfrac{1}{2}(1+g^2(s)). \tag{4}$$

This equation has the solution [recall that $g(0) = 1$]

$$g(s) = \tan(s/2 + \pi/4) = \tan s + 1/\cos s. \tag{5}$$

The power expansions of these functions are well known,[5] and we conclude that

$$\mu_{2k+1} = |E_{2k}|,$$
$$\mu_{2k} = 2^{2k}(2^{2k}-1)|B_{2k}|/(2k), \tag{6}$$

where E_{2k} and B_{2k} are Euler and Bernoulli numbers, respectively. The values for μ_N found here are larger than previously thought.[3,4,6] The first few are given in Table I.

As an illustration to our result, we list an element from each of the five classes in the case $N = 5$:

(1234) (5)	(1234) (5)	
(123) (4) (5)	(12) (34) (5)	
(12) (3) (4) (5)	(12) (3) (4) (5)	
(1) (2) (3) (4) (5)	(1) (2) (3) (4) (5)	
(123) (45)	(123) (45)	(123) (45)
(123) (4) (5)	(12) (3) (45)	(12) (3) (45)
(12) (3) (4) (5)	(12) (3) (4) (5)	(1) (2) (3) (45)
(1) (2) (3) (4) (5)	(1) (2) (3) (4) (5)	(1) (2) (3) (4) (5)

None of these chains can be obtained from another through a permutation.

It is also of interest to know the number of elements in the equivalence classes. Suppose that in a particular chain only one *pair* is broken, like in the first of the shown $N = 5$ chains. Permutations of the numbers in the pair leaves this chain invariant, but all other permutations generate a new chain belonging to the same class. The number of elements in the class is therefore $N!/2$. In other chains, pairs may be broken at two instances, like in all the other $N = 5$ cases. A

TABLE I. Number of coupled Yakubovskii-type equations for N identical particles.

N	3	4	5	6	7	8	9
μ_N	1	2	5	16	61	272	1385

[a)]Work supported by the Swedish Natural Science Research Council.

 0022-2488/82/030421-02$02.50 © 1982 American Institute of Physics

permutation of the numbers of either pair leaves the chain invariant, and the number of chains in each class is therefore $N!/2^2$ (at this point we can verify that our total number of 5-body chains is $5!/2 + 4\times 5!/2^2 = 180$, i.e., in accordance with the formula of Ref. 3).

In the general case, let $\mu_N^{(M)}$ be the number of N-body classes in which M pairs are broken in each chain, $\mu_N = \Sigma_M \mu_N^{(M)}$, $M \leqslant N/2$. Equation (1) generalizes to

$$\mu_N^{(M)} = \frac{1}{2}\sum_{k=1}^{N-1}\binom{N-2}{k-1}\sum_{i=0}^{M}\mu_k^{(i)}\mu_{N-k}^{(M-i)}, \tag{1'}$$

where $\mu_N^{(M)} = 0$, $M > N/2$. Obviously, $\mu_1^{(0)} = 1$, $\mu_2^{(0)} = 0$, $\mu_2^{(1)} = 1$. Let

$$\mu_N(t) = \sum_{i=0}^{\infty}\mu_N^{(i)}t^i, \tag{7}$$

Then (2) can be defined and (3) is again valid with μ_N understood as a function of t; in particular, $\mu_1 = 1$ and $\mu_2 = t$. Equation (4) takes the generalized form

$$\frac{d}{ds}g(s) = \tfrac{1}{2}(2t - 1 + g^2(s)) \tag{4'}$$

with the solution

$$g(s) = (2t-1)^{1/2}\tan((2t-1)^{1/2}s/2 + \arctan(2t-1)^{-1/2}). \tag{5'}$$

From the power series expansion of this expression (there is no singularity at $t = \frac{1}{2}$),

$$g(s) = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty}\mu_{N+1}^{(M)}t^M s^N/N!, \tag{2'}$$

the numbers $\mu_N^{(M)}$ are obtained.

As an illustration, consider the $N = 6$ case. From (5') we find that $\mu_6^{(0)} = 0$, $\mu_6^{(1)} = 1$, $\mu_6^{(2)} = 11$, and $\mu_6^{(3)} = 4$. There are therefore one class with $6!/2$ elements, 11 classes with $6!/2^2$ elements, and four classes with $6!/2^3$ elements. Adding up we find 2700 6-body chains, as expected.

In conclusion we have studied the set of complete chains of partitions of N objects, and in particular the equivalence classes generated by permutations. The number μ_N of such classes has been determined, as well as the number of elements in a given class, $N!/2^M$, where M is an easily found positive integer characteristic for the class. We have also determined the number $\mu_N^{(M)}$ of classes having the same number of elements (same M). Combining these results with a formula from Ref. 3, the total number of complete chains is

$$\sum_M \mu_N^{(M)}\frac{N!}{2^M} = N!(N-1)!/2^{N-1}. \tag{8}$$

This expression is a useful check on the calculation of the numbers $\mu_N^{(M)}$.

These combinatoric results are relevant for the formulation of the Yakubovskii-type N-body (quantum mechanical scattering) theory in the case of identical particles.

[1] O. A. Yakubovskii, Sov. J. Nucl. Phys. **5**, 937 (1967).
[2] B. R. Karlsson and E. M. Zeiger, Phys. Rev. D **9**, 1761 (1974); **10**, 1291 (1974).
[3] Gy. Bencze, Phys. Lett. B **72**, 155 (1977).
[4] Gy. Bencze and E. F. Redish, J. Math. Phys. **19**, 1909 (1978).
[5] M. Abramowitz and I. Stegun (eds.), *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1964), p. 75.
[6] L. P. Kok, Nucl. Phys. A **353**, 171c (1981).

A generalization of the Dirac equation to accelerating reference frames

John R. Urani
Physics Department, University of Missouri, Kansas City, Missouri 64110

Marilyn H. Kemp
Mathematics Department, University of Missouri, Kansas City, Missouri 64110

(Received 21 July 1981; accepted for publication 25 September 1981)

Using a recently developed global isometry method for treating accelerating observers, the induced tangent space transformation on flat Lorentzian R^4 is mapped homomorphically onto a time-dependent $D^{(1/2,0)} \oplus D^{(0,1/2)}$ representation of SL(2,C). The Dirac equation is shown to take on pseudoterms via this mapping. Eliminating the pseudoterms by identifying an affine connection, an exact analytic expression for the covariant derivative is found for general cases of arbitrary C^2 timelike observers. The transformation properties of the connection are shown to satisfy the conditions imposed by a general tetrad formalism. The specific case of the rotating observer is considered wherein the exact expression for the boosted Dirac equation is found.

PACS numbers: 04.20. – q

I. INTRODUCTION

Recently,[1-3] several exact treatments of accelerating observers in Lorentzian R^4 based on presymmetry[4] arguments have been published. Extensions of the above work are also available[5,6] treating the accelerating observer on an arbitrary space–time manifold.

An alternative[7,8] approach appeared recently utilizing global causality arguments along with Galilean and special relativistic limits to obtain a specific global isometric mapping on flat R^4 which induces a tangent space mapping consisting of time-dependent Lorentz boosts. Elimination of pseudoterms from particle and Maxwell equations allowed an affine connection to be identified for arbitrary C^2 timelike observers.

The standard 1→2 homomorphism of the full Lorentz group L_F onto SL(2,C) is used in this paper to consider the case of the Dirac equation relative to an accelerating observer in flat R^4. Straightforward manipulation of the derivative pseudoterm in the boosted Dirac equation leads to an exact expression for the affine connection which defines covariant differentiation. The connection obtained satisfies the general transformation conditions implied by the tetrad formalism of general relativity.[9]

The specific case of the rotating observer is considered where the Dirac equation relative to that noninertial observer is shown to take on acceleration and Thomas precession terms through the connection.

II. THE TANGENT SPACE HOMOMORPHISM

In Ref. 8. it was shown that a global isometry for R^4 could be found, parametrized by arbitrary interframe timelike velocity functions, which induced tangent-space maps consisting of time-dependent boosts. The Dirac equation for the field $\psi: R^4 \to C^2 \oplus C^2$ can be written in the covariant form (with bispinor source term R) as

$$-i\gamma^\mu \partial_\mu \psi + (mc/\hbar)\psi = R \tag{2.1}$$

relative to all inertial frames. The γ^μ matrices being invariant, it is obvious that any tangent space mapping of ∂_μ must be accompanied by an appropriate mapping L_D on $C^2 \oplus C^2$ in order to preserve the form of Eq. (2.1). This, of course, is the usual SL(2,C) direct-sum representation

$$L_D = D^{(1/2,0)} \oplus D^{(0,1/2)}. \tag{2.2}$$

The rigorous extension of Lorentz transformations to time-dependent boosts from Ref. 8 implies a mapping of the 4-gradient, $\partial'_\mu = \Lambda^{-1\alpha}{}_\mu(\beta(t))\partial_\alpha$, and suggests the extension of L_D of Eq. (2.2) via the use of the variable rapidity, $\beta(t)$, of the accelerating observer. It is also useful to extend the observer's trajectory, $X(\tau)$, to a congruence of curves on R^4 by simply translating the curve orthogonal to the time axis, thereby filling R^4 and defining a timelike C^1 vector field of tangents on R^4.

With metric $\eta = (+ - - -)$, we use the $D^{(1/2,0)}$ representation of the Lorentz boost,[10]

$$D^{(1/2,0)}(\beta(t)) = \cosh(\alpha/2) + \sinh(\alpha/2)(\sigma\cdot\hat{n}), \tag{2.3}$$

where $\cosh\alpha = \gamma(t)$, $\sinh\alpha = -\beta(t)\gamma(t)$, $\hat{n} = \beta(t)/\beta(t)$, and σ represents the Pauli matrices. Here, $\gamma(t) = (1-\beta^2(t))^{-1/2}$, as usual. The use of time dependent $\beta(t)$ does not alter the invariance properties of the Dirac matrices, i.e.,

$$\gamma^{\mu'} = \Lambda^\mu{}_\nu L_D \gamma^\nu L_D{}^{-1} = \gamma^\mu. \tag{2.4}$$

Consequently, writing the "boosted" Dirac field as $\psi' = L_D\psi$, etc., the left side of the Dirac equation becomes $-i\gamma^{\mu'}\partial'_\mu(L_D\psi) + (mc/\hbar)(L_D\psi)$. Using Eq. (2.4) we may express this as

$$\begin{aligned} &L_D(-i\gamma^\mu\partial_\mu\psi + (mc/\hbar)\psi) - i\gamma^{\mu'}(\partial'_\mu L_D)\psi \\ &\quad = L_D R - i\gamma^{\mu'}(\partial'_\mu L_D)\psi. \end{aligned} \tag{2.5}$$

Writing $\psi = L_D{}^{-1}\psi'$ and using $\partial'_\mu(L_D L_D{}^{-1}) = 0$ we may rewrite Eq. (2.5) in the covariant form

$$-i\gamma^{\mu'}\nabla'_\mu\psi' + (mc/\hbar)\psi' = R', \tag{2.6}$$

where $\nabla'_\mu = \partial'_\mu + \Gamma'_\mu(x')$ and

$$\Gamma'_\mu(x') = L_D(\partial'_\mu L_D{}^{-1}) \tag{2.7}$$

is the acceleration connection for $C^2 \oplus C^2$ induced by the accelerating observer. Obviously, for inertial frames, for

 0022-2488/82/030423-02$02.50 © 1982 American Institute of Physics

which $\beta = \text{const}$, Γ_μ is zero as expected. It is to be noted that the above connection has one upper and one lower spinor index suppressed and that it is similar in form to the connection for the tangent space itself.[8]

III. TRANSFORMATION OF Γ_μ

Successive induced tangent space mappings easily define the transformation properties of the above connection. Let O be an inertial frame, let O' be noninertial relative to O with rapidity $\beta(t)$ and let O'' be noninertial relative to O' with rapidity $\beta'(t')$. Composing the two spinor space mappings $L_D(\beta(t))$ and $L_D(\beta'(t'))$, the connection for O'' is

$$\Gamma''_\mu = (L_D(t')L_D(t))\partial''_\mu(L_D(t')L_D(t))^{-1}. \tag{3.1}$$

Using $\partial''_\mu = \Lambda^{-1}{}^{\alpha}{}_{\mu}(t')\partial'_\alpha$ we obtain, from Eq. (3.1),

$$\Gamma''_\mu = L_D(t')\partial''_\mu L_D{}^{-1}(t') + \Lambda^{-1}{}^{\alpha}{}_{\mu}(t')(L_D(t')\Gamma'_\alpha L_D{}^{-1}(t')), \tag{3.2}$$

where $\Gamma'_\alpha = L_D(t)\partial'_\alpha L_D{}^{-1}(t)$ is the connection of O'. This result is consistent with the general requirements for a connection[11] and is also consistent with the tetrad formalism[9] on an arbitrary space–time manifold. Here the tetrad is $\delta^{\alpha}{}_{\beta}$, for flat R^4, and it does not enter the connection since it is constant.

IV. APPLICATIONS

An application of the above formalism is to express the Dirac equation relative to a rotating observer where a Thomas precession would be expected.

For simplicity we take $\beta(t)$ to lie in the x-y plane and let $d\beta/dt = 0$. Then reversing the $\beta(t)$ factor in Eq. (2.3) we obtain the representation $D^{(0,1/2)}(\beta(t))$ and forming $L_D = D^{(1/2,0)} \oplus D^{(0,1/2)}$ we obtain connections using Eq. (2.7) in the form

$$\Gamma'_\mu = L_D\Lambda^{-1}{}^{0}{}_{\mu}(t)\left(\frac{\partial}{\partial ct}L_D{}^{-1}\right). \tag{4.1}$$

Choosing a fairly common representation[12] for the Dirac γ matrices, the Dirac equation relative to the rotating observer is, after lengthy calculation,

$$-i\left[\gamma^0\left(\frac{\partial}{\partial ct'} - \frac{-i}{2c}\mathbf{S}\cdot\boldsymbol{\omega}_T\right) + \boldsymbol{\gamma}\cdot\left(\nabla'_r + \frac{1}{2c}\frac{d\boldsymbol{\beta}}{d\tau}\right)\right]\psi' + \frac{mc}{\hbar}\psi' = R', \tag{4.2}$$

where $\boldsymbol{\omega}_T = (-(\gamma-1)/\beta^2)(\boldsymbol{\beta}\times d\boldsymbol{\beta}/dt)$ is the Thomas precession angular velocity, $\mathbf{S}$ is the spin operator whose z component is given by $S_z = \sigma^{12} = -i/2[\gamma^1,\gamma^2]$, $\psi' = L_D\psi$, and $d\boldsymbol{\beta}/d\tau = \gamma d\boldsymbol{\beta}/dt$. It is seen in Eq. (4.2) that the acceleration connection merely adds a spin interaction term and a radial centrifugal acceleration term relative to the rotating reference frame.

V. CONCLUSIONS

A straightforward extension of the Dirac equation to accelerating frames of reference in flat R^4 was accomplished via a mapping of time-dependent tangent space boosts onto a time-dependent $D^{(1/2,0)} \oplus D^{(0,1/2)}$ representation of SL(2,C). A closed form expression for the connection coefficients Γ_μ can be found by simple manipulation from Eq. (2.7). The acceleration-covariant Dirac equation [Eq. (2.6)] is completely equivalent to a tetrad formulation on flat R^4, where the tetrad may be taken to be the identity.

[1] B. DeFacio, P. W. Dennis, and D. G. Retzloff, Phys. Rev. D **18**, 2813 (1978).
[2] B. DeFacio, P. W. Dennis, and D. G. Retzloff, Phys. Rev. D **20**, 570 (1979).
[3] B. DeFacio and D. G. Retzloff, J. Math. Phys. **21**, 751 (1980).
[4] Y. Avishai and H. Ekstein, Commun. Math. Phys. **37**, 193 (1974).
[5] D. G. Retzloff, B. DeFacio, and P. W. Dennis, J. Math. Phys. **23**, 96 (1982).
[6] D. G. Retzloff, B. DeFacio, and P. W. Dennis, J. Math. Phys. **23**,105 (1982).
[7] J. R. Urani and G. D. Gale, Synthese (in press).
[8] J. R. Urani and M. H. Kemp, J. Math. Phys. **23**, 227 (1982).
[9] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 369.
[10] F. Halpern, *Special Relativity and Quantum Mechanics* (Prentice-Hall, Englewood Cliffs, N. J., 1968), p.11.
[11] F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965), p.288.
[12] S. Schweber, *Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), p.79.

The stationary coordinate systems in flat spacetime

John R. Letaw
Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712

Jonathan D. Pfautsch
Center for Theoretical Physics, The University of Texas at Austin, Austin, Texas 78712

(Received 2 December 1980; accepted for publication 9 January 1981)

The stationary metrics in flat spacetime are derived using a recent classification of the timelike Killing vector field trajectories. The metrics fall into six classes. A "simple" coordinate system from each class is selected as representative. Three of these systems are rectangular Minkowski coordinates, pseudocylindrical ("accelerating") coordinates, and rotating coordinates. The remaining three appear to be new coordinate types which will be useful in exploring coordinate-dependent effects in quantum field theory.

PACS numbers: 04.20.Cv

I. INTRODUCTION

Recently[1] the stationary motions in flat spacetime have been described and classified. These motions may be defined in at least two ways: (i) They are the world lines whose curvature invariants—curvature (κ), torsion (τ), and hypertorsion (υ)—are constant. Physically these invariants are the proper acceleration and angular velocity of the world line. Consequently, points on the world line are not distinguishable by local measurements. (ii) They are the integral curves, or trajectories, of the timelike Killing vector fields in flat spacetime.

The stationary motions fall into six classes according to the invariants. For a given choice of invariants, the world line is fixed up to an inhomogeneous Lorentz transformation. The classes are listed below with an expression for the tangent vectors in a convenient frame and a description in that frame. A frame-independent pictorial description of these motions is not possible, suggesting that they are best described locally by the frame-independent invariants. We denote proper time along a world line by s and use units where the speed of light equals 1 exclusively.

Class A:

$$\kappa = \tau = \upsilon = 0, \quad \dot{x}^\mu = (1,0,0,0). \tag{1}$$

This class contains inertial world lines.

Class B:

$$\tau = \upsilon = 0, \quad \dot{x}^\mu = (\cosh\kappa s, \sinh\kappa s, 0, 0). \tag{2}$$

This class consists of world lines along which the proper acceleration remains constant and equal to κ. The world lines are hyperbolic in this system.

Class C:

$$|\kappa| < |\tau|, \quad \upsilon = 0, \quad \rho^2 = \tau^2 - \kappa^2, \tag{3}$$
$$\dot{x}^\mu = \rho^{-1}(\tau, -\kappa\sin\rho s, \kappa\cos\rho s, 0).$$

In this frame the spatial projection of these world lines is a circle with a radius κ/ρ^2. The angular velocity is a constant ρ^2/τ.

Class D:

$$|\kappa| = |\tau|, \quad \upsilon = 0, \quad \dot{x}^\mu = (1 + \tfrac{1}{2}\kappa^2 s^2, \kappa s, \tfrac{1}{2}\kappa^2 s^2, 0). \tag{4}$$

The spatial projection is a semicubical parabola containing a cusp where the direction of motion is reversed. The cusp has no intrinsic distinction on the world line.

Class E:

$$|\kappa| > |\tau|, \quad \upsilon = 0, \quad \sigma^2 = \kappa^2 - \tau^2, \tag{5}$$
$$\dot{x}^\mu = \sigma^{-1}(\kappa\cosh\sigma s, \kappa\sinh\sigma s, \tau, 0).$$

The spatial projection of this motion is a catenary. A charged particle injected into a constant electric field follows a world line of this class if its initial velocity is perpendicular to the field. This is therefore the relativistic generalization of projectile motion. In the nonrelativistic ($s \ll \sigma^{-1}$, $s \ll \kappa^{-1}$) limit classes B, D, and E merge.

Class F:

$$\upsilon \neq 0,$$
$$\dot{x}^\mu = \left(\frac{\Delta}{R}\cosh R_1 s, \frac{\Delta}{R}\sinh R_1 s, \frac{-\kappa\tau}{\Delta R}\sin R_2 s, \frac{\kappa\tau}{\Delta R}\cos R_2 s\right),$$
$$R^4 = (\kappa^2 + \tau^2 + \upsilon^2)^2 - 4\kappa^2\tau^2,$$
$$\Delta^2 = \tfrac{1}{2}(R^2 + \kappa^2 + \tau^2 + \upsilon^2),$$
$$R_1^2 = \tfrac{1}{2}(R^2 + \kappa^2 - \tau^2 - \upsilon^2),$$
$$R_2^2 = \tfrac{1}{2}(R^2 - \kappa^2 + \tau^2 + \upsilon^2). \tag{6}$$

In this frame, this motion may be viewed as a rotating worldline uniformly accelerated normal to its plane of rotation. The radial distance from the x axis is maintained at a constant $\kappa\tau/\Delta R R_2$.

There are stationary coordinate systems adapted to each of these world lines because they are timelike Killing vector field trajectories. (By stationary we mean the metric is time-independent. This is to be distinguished from static which is a more specific term.) In these systems the lines of constant spatial coordinates, if timelike, are stationary world lines. The time coordinate is proportional to proper time. Each section of this paper is devoted to finding and describing the stationary coordinate systems adapted to a class of stationary motions. First, the Killing vector field associated with the class is constructed. This field and the requirement of stationarity determine the metric depending on four arbitrary functions of the spatial coordinates and an arbitrary constant. The arbitrary terms represent freedom to choose the spatial coordinates, the constant time hypersurfaces, and the time scale.

In addition to the general stationary coordinate system associated with each stationary motion, a particular system

 0022-2488/82/030425-07$02.50 © 1982 American Institute of Physics

which we feel is the most natural for each set of motions is exhibited. No attempt will be made to precisely define the criteria for "naturality"; our considerations are as follows. First, we seek the system with the maximum number of cyclic (not appearing in the metric) coordinates. Such a criterion distinguishes rectangular from spherical coordinates. Additionally, the dependence of the metric on noncyclic coordinates is minimized. For our purposes,[2] separability of the wave equation is another desideratum. Finally, the coordinate system associated with a more complex stationary motion should have as a limiting case that of a less complex motion.

The stationary coordinate systems find use in the study of the coordinate dependence of quantum field theory. They are particularly suited for this purpose because the metric is time-independent, allowing a globally conserved energy to be defined on constant time hypersurfaces. We discuss the quantization of the scalar field in the stationary coordinate systems in another paper.[2] Two of the coordinate systems developed here have an event horizon and a distinct stationary limit surface making them analogs of the rotating black hole metrics. Certain properties of these metrics might be elucidated by pursuing this analogy.

II. MINKOWSKI COORDINATES

The simplest of the stationary coordinate systems is the usual rectangular Minkowski coordinates. This system is defined by

Rectangular Minkowski coordinates:

$$x^0 = t, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z,$$

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \equiv \eta_{\mu\nu} dx^\mu dx^\nu.$$

These are not the only coordinates adapted to inertial world lines in the sense described in the Introduction. Any coordinate system in three-dimensional Euclidean space may be extended into a Minkowski-type system by the addition of an orthogonal time coordinate. In all these systems, the inertial world lines (1) are maintained at one spatial point, and proper time on these world lines is proportional to the temporal coordinate. The general stationary coordinate system adapted to the inertial world lines is developed by the method which follows.

The Killing vector field

$$\xi^\mu(x) = (1,0,0,0) \tag{7}$$

is tangent to all the world lines (1). All coordinate systems $(\bar{x}^0,\bar{x}^1,\bar{x}^2,\bar{x}^3)$ are sought in which this Killing vector field takes the form

$$\bar{\xi}^\mu(\bar{x}) = \chi^{-1}(1,0,0,0), \tag{8}$$

where χ is constant. In these systems the tangent vector to the inertial world lines necessarily has only a time component, and therefore the world lines are at rest.

The transformation between the Killing vector field (8) in barred coordinates and the field (7) in rectangular Minkowski coordinates is

$$\xi_\mu(x(\bar{x})) = \frac{\partial x^\mu(\bar{x})}{\partial \bar{x}^\nu} \bar{\xi}^\nu(\bar{x}). \tag{9}$$

This is a set of differential equations for the Minkowski coordinates in terms of the barred coordinates

$$(1,0,0,0) = \chi^{-1}\left(\frac{\partial t}{\partial \bar{x}^0}, \frac{\partial x}{\partial \bar{x}^0}, \frac{\partial y}{\partial \bar{x}^0}, \frac{\partial z}{\partial \bar{x}^0}\right). \tag{10}$$

These equations have as solutions the transformations

$$x^\mu = \chi \bar{x}^0 \delta^\mu_0 + A^\mu(\bar{x}^1,\bar{x}^2,\bar{x}^3), \tag{11}$$

where the A^μ are arbitrary functions of the spatial coordinates. The ease with which Eqs. (10) are solved may be traced to the lack of $\bar{x}$ dependence in $\bar{\xi}^\nu$.

Equation (11) generates coordinate systems adapted to the Killing vector field (7). The metric in these coordinates is

$$ds^2 = \chi^2(d\bar{x}^0)^2 + 2\chi A^0{}_{,\mu} d\bar{x}^0 d\bar{x}^\mu + \eta_{\alpha\beta} A^\alpha{}_{,\mu} A^\beta{}_{,\nu} d\bar{x}^\mu d\bar{x}^\nu. \tag{12}$$

Derivatives are with respect to the barred coordinates. The coordinate systems defined by (12) are the most general set of stationary coordinate systems associated with world lines of class A.

Among the class A stationary coordinate systems there are two which we wish to single out. The first is simply rectangular Minkowski coordinates. These are recovered from (12) by setting $\chi = 1$ and $A^\mu = \bar{x}^1\delta^\mu_1 + \bar{x}^2\delta^\mu_2 + \bar{x}^3\delta^\mu_3$. We consider these coordinates to be the most natural for a class A observer. The other system is cylindrical Minkowski coordinates defined by (12) with $\chi = 1$ and $A^\mu = \bar{x}^1\cos\bar{x}^2\delta^\mu_1 + \bar{x}^1\sin\bar{x}^2\delta^\mu_2 + \bar{x}^3\delta^\mu_3$. Thus we have

Cylindrical Minkowski coordinates:

$$\bar{x}^0 = t, \quad \bar{x}^1 = r, \quad \bar{x}^2 = \phi, \quad \bar{x}^3 = z,$$

$$r = (x^2 + y^2)^{1/2}, \quad \phi = \tan^{-1}(y/x),$$

$$ds^2 = dt^2 - dr^2 - r^2 d\phi^2 - dz^2.$$

In this and later coordinate descriptions new coordinates are defined when they appear. The definition is held to throughout the paper.

Rectangular and cylindrical Minkowski coordinates, as well as all Minkowski-type coordinates, are static. Static metrics, in addition to being time-independent, have no space-/time mixing components, that is, $g_{01} = g_{02} = g_{03} = 0$. In static coordinate systems the vector $\partial/\partial\bar{x}^0$ is orthogonal to the hypersurfaces $\bar{x}^0 = \text{const}$. The general stationary metric (12) associated with inertial motion is, however, not static. The nonstatic coordinate systems differ from Minkowski-type coordinates by having curved hypersurfaces of constant time.

In the remaining sections, we follow the methods of this section to generate the stationary coordinate systems adapted to each of the stationary motions.

III. PSEUDOCYLINDRICAL COORDINATES

The Killing vector field

$$\xi^\mu(x) = (\kappa x, \kappa t, 0, 0) \tag{13}$$

has as trajectories the world lines defined by (2). The coordinate systems associated with this field are given by the differential equations

$$(\kappa x, \kappa t, 0, 0) = \chi^{-1}\left(\frac{\partial t}{\partial \bar{x}^0}, \frac{\partial x}{\partial \bar{x}^0}, \frac{\partial y}{\partial \bar{x}^0}, \frac{\partial z}{\partial \bar{x}^0}\right). \tag{14}$$

The solutions of (14) are

$$x^\mu = A^\nu[(\delta^0_\nu\delta^\mu_0 + \delta^1_\nu\delta^\mu_1)\cosh\kappa\chi\bar{x}^0 + (\delta^0_\nu\delta^\mu_1 + \delta^1_\nu\delta^\mu_0)\sinh\kappa\chi\bar{x}^0 + (\delta^2_\nu\delta^\mu_2 + \delta^3_\nu\delta^\mu_3)], \quad (15)$$

where the A^ν are four arbitrary functions of the spatial variables.

The metric in the coordinate systems generated by (15) is

$$ds^2 = \kappa^2\chi^2 A^\alpha A^\beta(\delta^1_\alpha\delta^1_\beta - \delta^0_\alpha\delta^0_\beta)(d\bar{x}^0)^2 + 2\kappa\chi A^\alpha A^\beta{}_{,\mu} \times(\delta^1_\alpha\delta^0_\beta - \delta^0_\alpha\delta^1_\beta)d\bar{x}^0 d\bar{x}^\mu + \eta_{\alpha\beta}A^\alpha{}_{,\mu}A^\beta{}_{,\nu}d\bar{x}^\mu d\bar{x}^\nu. \quad (16)$$

This metric represents the most general stationary coordinate system adapted to uniformly accelerated world lines.

If $\chi = \kappa^{-1}$ and $A^\nu = \bar{x}^1\delta^\nu_1 + \bar{x}^2\delta^\nu_2 + \bar{x}^3\delta^\nu_3$, a familiar coordinate system is recovered.

Pseudocylindrical coordinates:

$$\bar{x}^0 = \tau, \quad \bar{x}^1 = \xi, \quad \bar{x}^2 = y, \quad \bar{x}^3 = z,$$

$$ds^2 = \xi^2 d\tau^2 - d\xi^2 - dy^2 - dz^2,$$

$$\tau = \tanh^{-1}(t/x), \quad \xi = (t^2 - x^2)^{1/2}.$$

In this system the metric is static and there are three cyclic coordinates. This system has been discussed in detail by Rindler[3] and has been used extensively because of its similarity to Kruskal[4] coordinates. It has been called accelerating, hyperbolic, Fermi, and Rindler coordinates by various authors. The name "accelerating coordinates" does not communicate the static nature of the metric, while "hyperbolic coordinates" does not suggest space/time mixing and is subject to confusion with three-dimensional Euclidean systems. We therefore propose the name "pseudocylindrical coordinates" because it better describes the inherent geometrical properties of the system. These coordinates are the natural generalization of cylindrical coordinates to a space with indefinite metric. Thus a cylinder in Euclidean space becomes a pseudocylinder in pseudo-Euclidean space. This name is subject to some confusion because an acceleration is sometimes called a pseudorotation. If it is recalled that coordinate systems are generally named after coordinate surfaces, not coordinate lines, this confusion will be minimized.

Some aspects of the pseudocylindrical coordinate system are displayed in Fig. 1. As they are expressed above, pseudocylindrical coordinates cover only one-quarter of spacetime (region I in the figure); however, slight changes in the definitions of τ and ξ yield suitable coordinates for the remaining regions. For example, ξ might be defined as $-(t^2 - x^2)^{1/2}$ in region III while τ is defined as in region I. The resulting metric is stationary only in regions I and III, where it is also static.

Taking region I to be the physical region, one finds that the surface $\tau = -\infty, \xi = 0$ is a past event horizon and the surfaces $\tau = \infty, \xi = 0$ is a future event horizon. In general the term *event horizon* is accurately applied only with reference to one world line.[5] That is, it is the boundary of spacetime regions which are in some manner causally disconnected from the world line. In special cases, the timelike coordinate lines in one region have, when viewed as world lines, identical event horizons. When this occurs, one may

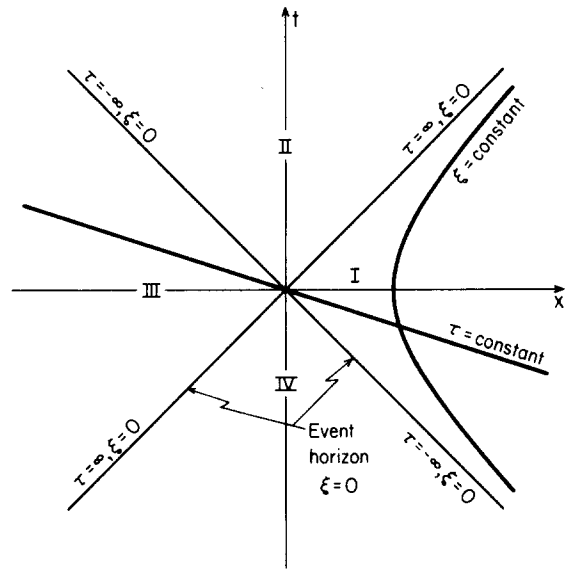

FIG. 1. Pseudocylindrical coordinate surfaces displayed in rectangular Minkowski coordinates. Spacetime is divided into four regions (I, II, III, IV) by the event horizons $\tau = \infty, \xi = 0$ and $\tau = -\infty, \xi = 0$. An object on a ξ = constant line has proper acceleration ξ^{-1}. Proper time on this world line is $s = \xi\tau$.

speak of the event horizon as associated with the coordinate system because a spatial region is causally disconnected at all times in the same manner as the world lines. Such an event horizon is indicated whenever there is a null surface which is also a coordinate surface, and is obviously coordinate-dependent. A notable example of this type of event horizon is found in Schwarzschild coordinates,[4] The event horizons of pseudocylindrical coordinates, and the other systems in this paper are also of this type. In the general system, event horizons associated with individual coordinate lines are different, and the coordinate system cannot be said to have an event horizon. Several of the flat, conformally static systems of Brown, Ottewill, and Siklos[6] are pretty examples of this. Each coordinate line is hyperbolic (uniformly accelerated); yet there is no coordinate event horizon.

IV. ROTATING COORDINATES

World lines of class C are trajectories of the Killing vector field

$$\xi^\mu(x) = (\tau/\rho, -\rho y, \rho x, 0). \quad (17)$$

Coordinate systems associated with this field are found to be given by the transformations

$$x^\mu = A^\nu[(\delta^0_\nu\delta^\mu_0 + \delta^3_\nu\delta^\mu_3) + (\delta^1_\nu\delta^\mu_1 + \delta^2_\nu\delta^\mu_2)\cos\rho\chi\bar{x}^0 + (\delta^1_\nu\delta^\mu_2 - \delta^2_\nu\delta^\mu_1)\sin\rho\chi\bar{x}^0] + \chi(\tau/\rho)\bar{x}^0\delta^\mu_0, \quad (18)$$

and the metric is

$$ds^2 = [\chi^2\tau^2/\rho^2 - \chi^2\rho^2 A^\alpha A^\beta(\delta^1_\alpha\delta^1_\beta + \delta^2_\alpha\delta^2_\beta)](d\bar{x}^0)^2 + 2\chi[(\tau/\rho)A^\alpha{}_{,\mu}\delta^0_\alpha - \rho A^\alpha A^\beta{}_{,\mu}(\delta^1_\alpha\delta^2_\beta - \delta^2_\alpha\delta^1_\beta)]d\bar{x}^0 d\bar{x}^\mu + \eta_{\alpha\beta}A^\alpha{}_{,\mu}A^\beta{}_{,\nu}d\bar{x}^\mu d\bar{x}^\nu. \quad (19)$$

There are no static forms of the metric (19) because there is no surface everywhere orthogonal to the Killing vec-

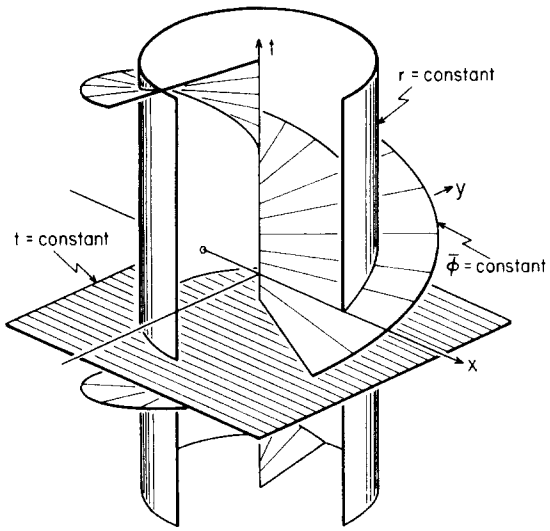

FIG. 2. Rotating coordinate surfaces displayed in rectangular Minkowski coordinates. Constant t surfaces are planes and constant r surfaces are circular cylinders identical to surfaces in cylindrical Minkowski coordinates. Constant $\bar{\phi}$ surfaces are "twisted" versions of constant ϕ surfaces resembling a spiral staircase or Archimedean screw.

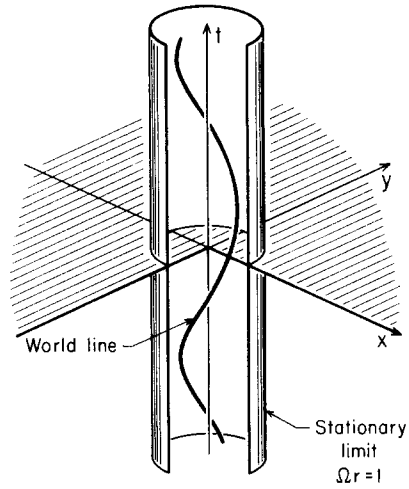

FIG. 3. The environment of a helical world line as it appears in rectangular Minkowski coordinates. An observer on the world line has a velocity of one-half the speed of light in this frame of reference. The stationary limit surface separates the timelike region of the Killing vector field from the spacelike part. All objects outside the stationary limit will appear to be "dragged" around clockwise relative to the world line.

tor field (17). Letting $\chi = \rho/\tau$ and $A^{\mu} = \bar{x}^1\cos\bar{x}^2\delta^{\mu}_1 + \bar{x}^1\sin\bar{x}^2\delta^{\mu}_2 + \bar{x}^3\delta^{\mu}_3$ gives the usual rotating coordinate system.[7]

Rotating coordinates:

$$\bar{x}^0 = t, \quad \bar{x}^1 = r, \quad \bar{x}^2 = \bar{\phi}, \quad \bar{x}^3 = z,$$
$$ds^2 = dt^2 - dr^2 - r^2(d\bar{\phi} + \Omega dt)^2 - dz^2,$$
$$\bar{\phi} = \phi - \Omega t.$$

This coordinate system depends on one free parameter $\Omega = \rho^2/\tau$ which must be constant. Because of this parameter, any rotating coordinate system is adapted only to one subclass of the world lines in class C. This is to be distinguished from pseudocylindrical coordinates where no free parameters occur, and all world lines of class B appear as coordinate lines.

The coordinate surfaces of rotating coordinates are shown in Fig. 2. Surfaces of constant r and t are identical to coordinate surfaces in cylindrical Minkowski coordinates. The difference between the systems lies in the arrangement of the coordinate lines on the cylindrical surfaces. Lines of constant $\bar{\phi}$ are displaced relative to lines of constant ϕ at a constant rate Ω (relative to coordinate time for both systems), thus forming helices rather than vertical straight lines. As a result of this displacement, surfaces of constant $\bar{\phi}$ are right conoids resembling the screw of Archimedes. The name "rotating coordinates" is well established. A suggestion of motion is desirable because the coordinates are not static, that is, motion along constant t hypersurfaces is implies because the Killing vectors are not orthogonal to it.

A particle at rest in rotating coordinates moves along a helical world line with angular velocity Ω and radius r as viewed from Minkowski coordinates (see Fig. 3). The particle has proper acceleration $\kappa = \Omega^2 r/(1 - \Omega^2 r^2)$ and angular velocity $\tau = \Omega/(1 - \Omega^2 r^2)$. The Killing vector field (17) becomes null on the surface $r = 1/\Omega$. Beyond this stationary limit all Killing vector trajectories are spacelike, thus no object can be at rest relative to the rotating coordinates. The explanation for this is that no object can move faster than the speed of light. The region is similar to the ergosphere surrounding a rotating black hole, although no energy can be extracted from it. There is no event horizon in rotating coordinates.

V. NULL PARABOLIC COORDINATES

The Killing vector field

$$\xi^{\mu}(x) = (1 + \kappa x, \kappa t - \kappa y, \kappa x, 0) \tag{20}$$

has the world lines of class D as trajectories. This form of the Killing vector cannot be significantly simplified by transformation to another Lorentz frame; that is, it cannot be represented as a linear combination of commuting boost, rotation, and translation Killing vectors as with the other motions.[8] Because of the complexity of the Killing vector field a coordinate system based on cylinders or pseudocylinders is not possible.

Transformations derived from the Killing vector field (20) are

$$\begin{aligned} x^{\mu} = {} & \tfrac{1}{6}\kappa^2\chi^3(\bar{x}^0)^3(\delta^{\mu}_0 + \delta^{\mu}_2) + \tfrac{1}{2}\kappa\chi^2\delta^{\mu}_1(\bar{x}^0)^2 + \chi\bar{x}^0\delta^{\mu}_0 \\ & + A^{\nu}\{\kappa^2\chi^2(\bar{x}^0)^2(\delta^0_{\nu}\delta^{\mu}_0 + \delta^0_{\nu}\delta^{\mu}_2) \\ & + \kappa\chi\bar{x}^0(\delta^0_{\nu}\delta^{\mu}_1 + \delta^1_{\nu}\delta^{\mu}_0 + \delta^1_{\nu}\delta^{\mu}_2) \\ & + (\delta^1_{\nu}\delta^{\mu}_1 + \delta^2_{\nu}\delta^{\mu}_0 - \delta^0_{\nu}\delta^{\mu}_2 + \delta^2_{\nu}\delta^{\mu}_2 + \delta^3_{\nu}\delta^{\mu}_3)\}, \end{aligned} \tag{21}$$

and the metrics associated with these transformations are

$$\begin{aligned} ds^2 = {} & \chi^2(1 + 2\kappa A^{\alpha}\delta^1_{\alpha} - \kappa^2 A^{\alpha}A^{\beta}\delta^0_{\alpha}\delta^0_{\beta})(d\bar{x}^0)^2 + 2\chi A^{\alpha}{}_{,\mu}\delta^2_{\alpha} \\ & \times d\bar{x}^0 d\bar{x}^{\mu} + [\eta_{\alpha\beta} + (\delta^2_{\alpha} - \delta^0_{\alpha})(\delta^2_{\beta} - \delta^0_{\beta}) \\ & - \delta^0_{\alpha}\delta^0_{\beta}]A^{\alpha}{}_{,\mu}A^{\beta}{}_{,\nu}d\bar{x}^{\mu}d\bar{x}^{\nu}. \end{aligned} \tag{22}$$

Some general features of the coordinate systems can be found. First, from the transformation (21) one finds

$$\chi\bar{x}^0 + A^0(\bar{x}^1, \bar{x}^2, \bar{x}^3) = t - y. \tag{23}$$

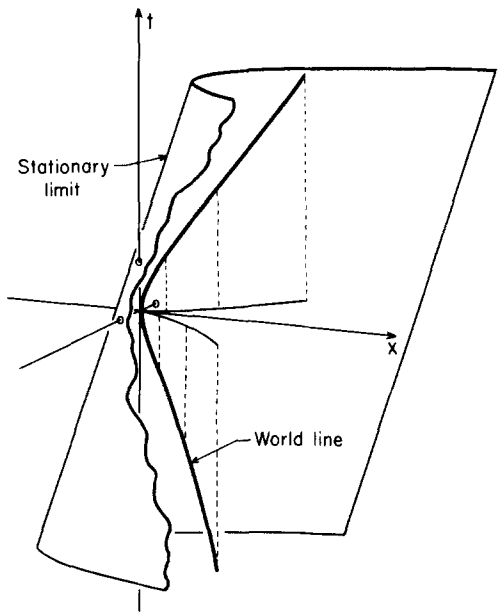

FIG. 4. The environment of a class D world line in rectangular Minkowski coordinates. The world line is defined by $\kappa = \tau$ and has as its spatial projection the semicubical parabola shown. An observer on the world line would measure constant proper acceleration equal to constant proper angular velocity. The world line is causally connected with all points in spacetime; therefore, there is no event horizon in a class D coordinate system. However, there is a stationary limit surface (parabolic cylinder in this frame), and a corresponding region where no world lines are at rest. The parabolic cylinder has a null axis which remains null in different Lorentz frames.

Thus, to set A^0 equal to zero necessitates $\bar{x}^0$ becoming a null coordinate. Second, the parameter κ must appear in the metric. Third, there is no event horizon in this coordinate system. This is easily verified by showing there are past- and future-directed null geodesics between each point in spacetime and some point on the world line whose tangent vector is (4). Finally, there is a stationary limit where (20) becomes null. The surface is a parabolic cylinder (with a null axis) defined by

$$1 + 2\kappa x - \kappa^2(t-y)^2 = 0. \tag{24}$$

This surface and a representative world line are shown in Fig. 4.

A coordinate system with a particularly simple metric is derived from (22) if $\chi = 1$ and

$$A^\mu = (\bar{x}^1 - 1/2\kappa)\delta^\mu_1 + \bar{x}^2\delta^\mu_2 + \bar{x}^3\delta^\mu_2. \tag{25}$$

Null parabolic coordinates:

$$ds^2 = 2\kappa\bar{x}^1(d\bar{x}^0)^2 + 2d\bar{x}^2d\bar{x}^0 - (d\bar{x}^1)^2 - (d\bar{x}^3)^2,$$
$$x^0 = \tfrac{1}{6}\kappa^2(\bar{x}^0)^3 + (\kappa\bar{x}^1 + \tfrac{1}{2})\bar{x}^0 + \bar{x}^2,$$
$$x^1 = \tfrac{1}{2}\kappa(\bar{x}^0)^2 + (\bar{x}^1 - \tfrac{1}{2}\kappa),$$
$$x^2 = \tfrac{1}{6}\kappa^2(\bar{x}^0)^3 + (\kappa\bar{x}^1 - \tfrac{1}{2})\bar{x}^0 + \bar{x}^2,$$
$$x^3 = \bar{x}^3.$$

The metric has three cyclic coordinates, and the wave equation is separable. This appears to be the most "natural" of the class D coordinate systems.

VI. PSEUDOROTATING COORDINATES

The class E motions are trajectories of the Killing vector field

$$\xi^\mu(x) = (\sigma x, \sigma t, \tau/\sigma, 0). \tag{26}$$

The transformations to stationary coordinate systems are therefore

$$x^\mu = A^\nu\{(\delta^0_\nu\delta^\mu_0 + \delta^1_\nu\delta^\mu_1)\cosh\sigma\chi\bar{x}^0 + (\delta^0_\nu\delta^\mu_1 + \delta^1_\nu\delta^\mu_0)\sinh\sigma\chi\bar{x}^0 + (\delta^2_\nu\delta^\mu_2 + \delta^3_\nu\delta^\mu_3)\} + \chi\frac{\tau}{\sigma}\bar{x}^0\delta^\mu_2, \tag{27}$$

and the metrics are

$$ds^2 = [\sigma^2\chi^2A^\alpha A^\beta(\delta^1_\alpha\delta^1_\beta - \delta^0_\alpha\delta^0_\beta) - \chi^2\tau^2/\sigma^2](d\bar{x}^0)^2 + 2[\chi(\tau/\sigma)A^\alpha{}_{,\mu}\delta^2_\alpha - \sigma\chi A^\alpha A^\beta{}_{,\mu}(\delta^0_\alpha\delta^1_\beta - \delta^1_\alpha\delta^0_\beta)]d\bar{x}^0d\bar{x}^\mu + \eta_{\alpha\beta}A^\alpha{}_{,\mu}A^\beta{}_{,\nu}d\bar{x}^\mu d\bar{x}^\nu. \tag{28}$$

Metrics associated with class E motions, like those associated with class C motions, are never static because there are no surfaces everywhere orthogonal to the vector field (26). The choice $\chi = 1/\sigma$ and $A^\mu = \bar{x}^1\delta^\mu_1 + \bar{x}^2\delta^\mu_2 + \bar{x}^3\delta^\mu_3$ leads to the metric for

Pseudorotating coordinates:

$$\bar{x}^0 = \tau, \quad \bar{x}^1 = \xi, \quad \bar{x}^2 = \bar{y}, \quad \bar{x}^3 = z,$$
$$ds^2 = \xi^2d\tau^2 - d\xi^2 - (d\bar{y} + \bar{\Omega}d\tau)^2 - dz^2,$$
$$\bar{y} = y - \bar{\Omega}\tau.$$

This system depends on the parameter $\bar{\Omega} = \tau/\sigma^2$ and has as coordinate lines only a subclass of world lines in class E.

Coordinate surfaces in the pseudorotating system are shown in Fig. 5. Surfaces of constant ξ and τ are identical to surfaces in pseudocylindrical coordinates. Because lines of constant $\bar{y}$ move at a uniform rate $\bar{\Omega}$ (relative to coordinate

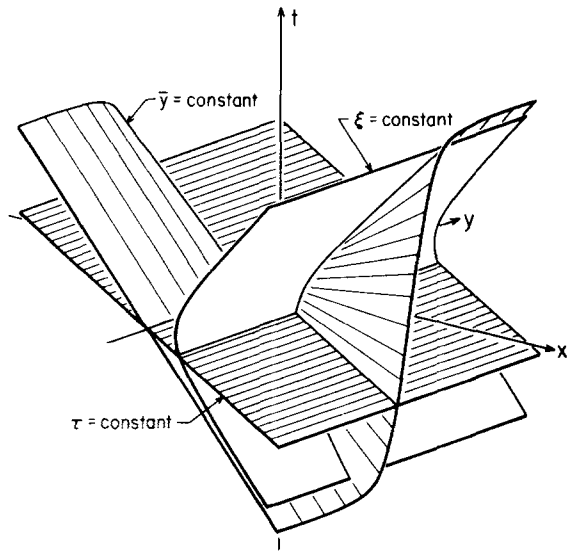

FIG. 5. Pseudorotating coordinate surfaces displayed in rectangular Minkowski coordinates. Constant τ surfaces are planes, and constant ξ surfaces are hyperbolic cylinders just as in pseudocylindrical coordinates. The constant $\bar{y}$ surface is a "twisted"version of the constant y plane. Note that the $\bar{y}$ = const surface and the τ = const surface intersect in a y = const plane and that this plane depends on both $\bar{y}$ and τ ($y = \bar{y} + \bar{\Omega}\tau$). Note also that surfaces of constant $\bar{y}$ and τ are not orthogonal.

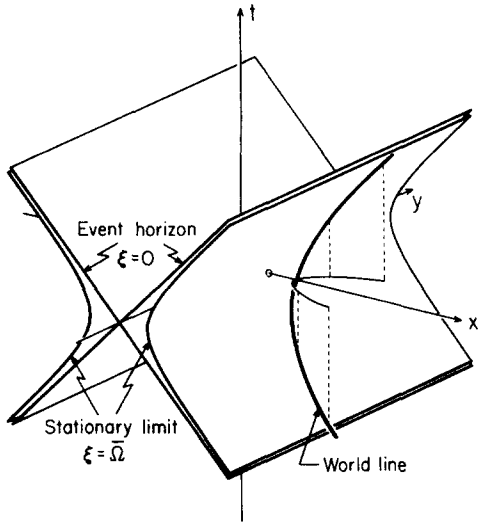

FIG. 6. Environment of a class E world line in rectangular Minkowski coordinates. A world line with $\kappa = 2\tau$ is shown, its spatial projection being a catenary in this frame. An observer on this world line measures constant proper acceleration equal to twice his proper angular velocity. The event horizon at $\xi = 0$ divides spacetime just as in pseudocylindrical coordinates. A stationary limit surface separates timelike from spacelike Killing vectors. In the region between the event horizon and the stationary limit no object can be at rest relative to pseudorotating coordinates.

time τ) relative to lines of constant y, the surfaces of constant $\bar{y}$ are twisted in a manner analogous to the twisting of $\tilde{\phi} = \text{const}$ surfaces in rotating coordinates. These surfaces are once again right conoids and, as required in a nonstatic system, are not orthogonal to surfaces of constant τ. We choose the name "pseudorotating" for this coordinate system because it resembles rotating coordinates without having any periodic motion and because this suggests its relation to pseudocylindrical coordinates.

The environment of a particle at rest in pseudorotating coordinates is displayed in Fig. 6. Event horizons divide the manifold into four regions exactly as in pseudocylindrical coordinates. In addition to the event horizons there is a stationary limit surface at $\xi = \bar{\Omega}$, where the Killing vector field (26) becomes null. Between the stationary limit and the event horizon is a region where all timelike world lines move relative to pseudorotating coordinates. This region is strongly suggestive of the ergosphere of a rotating black hole.

Pseudorotating coordinate lines are world lines of class E with proper acceleration $\kappa = \xi/(\xi^2 - \bar{\Omega}^2)$ and angular velocity $\tau = \bar{\Omega}/(\xi^2 - \bar{\Omega}^2)$. These world lines are physical only when $\xi^2 > \bar{\Omega}^2$.

Class E world lines should not be confused with the Lorentz transforms of class B world lines. The constant $\bar{y}$ surface is a function of τ, not t, and is therefore not a plane but a curved surface. Also, the spatial projection of a Lorentz transformed hyperbolic world line is necessarily a hyperbola, not a catenary as in Fig. 6. The y component of velocity relative to the Minkowski coordinates of this figure decreases with time $[dy/dt = \bar{\Omega}/(\xi^2 + t^2)^{1/2}]$.

A possible alternative to the pseudorotating coordinate system is arrived at by setting $\chi = \sigma/\tau$ and $A^\mu = \bar{x}^1 \sinh \bar{x}^2 \delta^\mu_0 + \bar{x}^1 \cosh \bar{x}^2 \delta^\mu_1 + \bar{x}^3 \delta^\mu_3$. The metric for this system is

$$ds^2 = -(d\bar{x}^0)^2 - (d\bar{x}^1)^2 + (\bar{x}^1)^2(d\bar{x}^2 + \bar{\Omega}^{-1} d\bar{x}^0)^2 - (d\bar{x}^3)^2. \tag{29}$$

Comparison of this metric with that of rotating coordinates explicitly shows the similarities between them. In spite of this, we prefer pseudorotating coordinates because they become pseudocylindrical coordinates in the limit $\bar{\Omega} \to 0$.

VII. ROTATING PSEUDOCYLINDRICAL COORDINATES

Motions of class F are the most general stationary world lines in the sense that all three curvature invariants are nonzero, and they cannot be described in a spacetime of less than four dimensions. These world lines are trajectories of the Killing vector field

$$\xi^\mu(x) = (R_1 x, R_1 t, -R_2 z, R_2 y). \tag{30}$$

Coordinate transformations derived from the field are

$$\begin{aligned} x^\mu = A^\nu [&(\delta^0_\nu \delta^\mu_0 + \delta^1_\nu \delta^\mu_1)\cosh R_1 \chi \bar{x}^0 + (\delta^0_\nu \delta^\mu_1 + \delta^1_\nu \delta^\mu_0) \\ &\times \sinh R_1 \chi \bar{x}^0 + (\delta^2_\nu \delta^\mu_2 + \delta^3_\nu \delta^\mu_3)\cos R_2 \chi \bar{x}^0 \\ &+ (\delta^2_\nu \delta^\mu_3 - \delta^3_\nu \delta^\mu_2)\sin R_2 \chi \bar{x}^0], \end{aligned} \tag{31}$$

and the metrics are

$$\begin{aligned} ds^2 = &(R_1^2 + R_2^2)\chi^2 \eta_{\alpha\beta} A^\alpha A^\beta (d\bar{x}^0)^2 \\ &- 2\chi (A^\alpha A^\beta{}_{,\mu} - A^\beta A^\alpha{}_{,\mu})(R_1 \delta^0_\alpha \delta^1_\beta + R_2 \delta^2_\alpha \delta^3_\beta) d\bar{x}^0 d\bar{x}^\mu \\ &+ \eta_{\alpha\beta} A^\alpha{}_{,\mu} A^\beta{}_{,\nu} d\bar{x}^\mu d\bar{x}^\nu. \end{aligned} \tag{32}$$

A particular coordinate system in this class of metrics is found by letting $\chi = 1/R_1$ and $A^\nu = \bar{x}^1 \delta^\nu_1 + \bar{x}^2 \cos \bar{x}^3 \delta^\nu_2 + \bar{x}^2 \sin \bar{x}^3 \delta^\nu_3$.

Rotating pseudocylindrical coordinates:

$$\bar{x}^0 = \tau, \quad \bar{x}^1 = \xi, \quad \bar{x}^2 = r, \quad \bar{x}^3 = \tilde{\phi},$$

$$ds^2 = \xi^2 d\tau^2 - d\xi^2 - dr^2 - r^2(d\tilde{\phi} + \tilde{\Omega} d\tau)^2,$$

$$r = (y^2 + z^2)^{1/2}, \quad \tilde{\phi} = \tan^{-1}(z/y) - \tilde{\Omega}\tau.$$

Some remarks concerning coordinate definitions are in order. The definition of r differs slightly from earlier usage, but it is essentially the same coordinate. This minor change might be avoided by requiring (contrary to convention) that all accelerations be in the z direction and all rotations be around the z axis. The similarity of the definition of $\tilde{\phi}$ and $\bar{\phi}$ is only formal because the rotations are measured relative to different coordinate times. The parameter $\tilde{\Omega} = R_2/R_1$ singles out one subclass of the world lines of class F.

"Rotating pseudocylindrical coordinates" is an appropriate name for this system because it is a pseudocylindrical coordinate system with an essentially independent rotation tacked on. A rotation of the Minkowski (y,z) plane relative to the Minkowski (t,x) plane does not upset coordinate relations in the (t,x) plane where the pseudocylindrical coordinates are defined. Surfaces of constant ξ and τ are exactly the same as the corresponding surfaces in pseudocylindrical and pseudorotating coordinates. Surfaces of constant r and $\tilde{\phi}$ appear the same as constant r and $\bar{\phi}$ surfaces in rotating coordinates (Fig. 2) if the Minkowski t axis is replaced with a τ axis. Relative to the Minkowski t axis the pitch of the screw increases rapidly with increasing t.

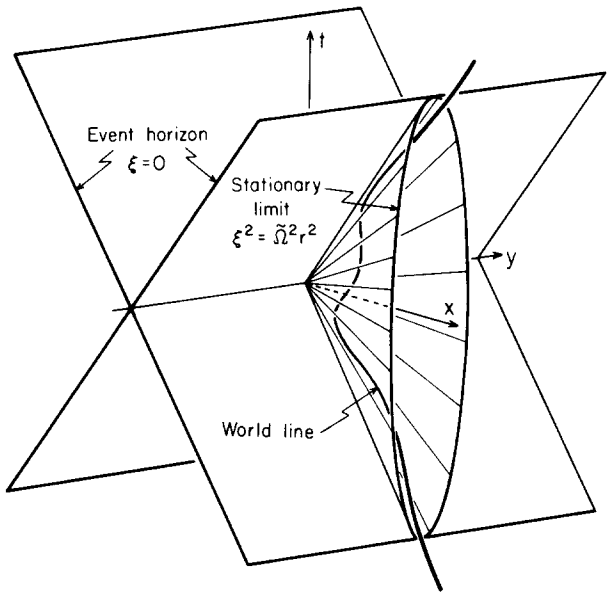

FIG. 7. The environment of a rotating and accelerating (class F) world line in rectangular Minkowski coordinates. The event horizon at $\xi = 0$ is exactly as in pseudocylindrical and pseudorotating coordinates. The stationary limit is an elliptical cone when the z axis is suppressed, but a more arbitrary slice $z = \text{const} \times y$ yields an elliptic hyperboloid. A world line moving at close to the speed of light is shown. Its rotation appears as an oscillation along the y axis; its acceleration appears as the hyperbolic form in the (x,t) plane. The oscillation is redshifted at large proper times.

The environment of a static observer in rotating pseudocylindrical coordinates is shown in Fig. 7 with one dimension suppressed. Event horizons occur exactly as in pseudocylindrical and pseudorotating coordinates. The Killing vector field (30) becomes null on the surface $\xi^2 = \tilde{\Omega}^2 r^2$ which can be various elliptic hyperboloids, depending on the exact choice of ϕ in the projection. In Fig. 7 the z axis is suppressed ($\phi = 0$) and the stationary limit appears as an elliptic cone. There are no static world lines outside this cone.

A representative stationary world line is also shown in Fig. 7. The world line has proper acceleration $\kappa = (\xi^2 + \tilde{\Omega}^4 r^2)^{1/2}/(\xi^2 - \tilde{\Omega}^2 r^2)$ and angular velocity $(\tau^2 + v^2)^{1/2} = \tilde{\Omega}(\xi^2 + r^2)^{1/2}/(\xi^2 - \tilde{\Omega}^2 r^2)$. As expected, these diverge on the stationary limit surface. When $\xi \gg r$ and $\tilde{\Omega} \ll 1$, we recover the uniformly accelerated observer. The apparent change in the frequency of oscillation in Fig. 7 is simply a redshift effect occurring because the velocity approaches the speed of light as proper time $s \to \infty$.

VIII. CONCLUSION

In this paper we have taken a complete set of timelike Killing vector fields in flat space and derived all stationary coordinate systems adapted to them. The fields are complete in the following sense: There is a threefold infinity of trajectories of Killing vector fields defined by κ, τ, and v and listed explicitly in Eqs. (1)–(6). Any timelike stationary world line is the Lorentz transform of one of these trajectories. For each of the trajectories, a Killing vector field coincident with its tangent vector is found. The resulting set of fields is therefore representative of all timelike Killing vector fields because any other can be generated from these by a Lorentz transformation. The metrics eventually derived from these Killing vector fields are the most general set of stationary metrics in flat space because they are unaffected by transformations of the fields.

A more elegant set of Killing vector fields from which the stationary coordinate systems might be derived is

$$\xi^\mu(x) = (1 + \kappa x, \kappa t - \tau y, \tau x - v z, v y). \tag{33}$$

Any timelike Killing vector field may be put in this form by a Lorentz transformation to the frame in which the world line is at rest when at the origin of coordinates. This representation is not as convenient for calculation as the representations we have chosen.

In general, each stationary metric is fixed only up to four arbitrary functions of the spatial coordinates and a constant. The constant is a scale factor on the time coordinate, and the arbitrary functions represent the freedom to vary spatial coordinates and constant time hypersurfaces without undoing the stationarity of the metric. We have singled out a particular case of each type of stationary metric for use in later work according to criteria listed in the Introduction. In all cases a satisfactory system for these purposes has been found.

ACKNOWLEDGMENTS

The authors wish to thank Duane A. Dicus and John A. Wheeler for their continuing support. This work was funded in part by grants from the Department of Energy and the National Science Foundation (PHY 7826592).

[1] J. R. Letaw, "Stationary world lines and the vacuum excitation of noninertial detectors," Phys. Rev. D **23**, 1709 (1981).
[2] J. R. Letaw and J. D. Pfautsch, "The quantized scalar field in stationary coordinate systems of flat spacetime," Phys. Rev. D **24**, 1491 (1981).
[3] W. Rindler, Am. J. Phys. **34**, 1174 (1966).
[4] M. D. Kruskal, Phys. Rev. **119**, 1743 (1960).
[5] W. Rindler, Mon. Not. R. Astron. Soc. **116**, 662 (1956).
[6] M. Brown, A. Ottewill, and S. Siklos (unpublished results).
[7] J. R. Letaw and J. D. Pfautsch, Phys. Rev. D **22**, 1345 (1980).
[8] Equation (20) cannot be put into simpler form for the same reason that equal and perpendicular electric and magnetic fields cannot be reduced to either a pure electric or a pure magnetic field. This is not entirely fortuitous since a charged particle injected into these fields will follow a class D world line.

Singular boundaries of space-times

Robert Geroch, Liang Can-bin, and Robert M. Wald
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

(Received 18 August 1981; accepted for publication 2 October 1981)

We give an example of a causally well-behaved, singular space–time for which all singular-boundary constructions which fall in a certain wide class—a class which includes both the *g*-boundary and *b*-boundary—yield pathological topological properties. Specifically, for such a construction as applied to this example, a singular boundary point fails to be T_1-related to an event of the original space–time. This example suggests that there may not exist any useful, generally applicable notion of the singular boundary of a space–time.

PACS numbers: 04.20.Cv

1. INTRODUCTION

It has been known for many years that large classes of solutions of Einstein's equation must be singular. However, it has proven difficult to formulate a precise description of this singular behavior, since space–times in the framework of general relativity do not include "singular events" at which a local analysis of structure can be carried out. This circumstance has led to the search for a construction which determines, for any singular space–time, some additional singular points attached to that space–time. In particular, a useful algorithm has been sought which assigns, to each space–time (M,g_{ab}), a topological space $\bar{M}$ with M as an open, dense topological subspace. The points of $\bar{M} - M$ are to be interpreted as the "singular points" of (M,g_{ab}), while the topology of $\bar{M}$ would describe "how close" regular space–time events are to singular points.

At least two such constructions for singular boundaries have been proposed: those for the *g*-boundary[1] and the *b*-boundary.[2] In the *g*-boundary construction, the singular points are introduced as ideal endpoints of incomplete geodesics, while the topology arises from the behavior of these geodesics under small variations of their initial conditions. In the *b*-boundary construction, the singular points are ideal endpoints of those curves incomplete with respect to a suitable generalization of affine parameter, while the topology arises from a certain "distance," also defined in terms of this parameter.

It is known that the *b*-boundary construction gives unphysical results even for common solutions of Einstein's equation. In particular, for the extended Schwarzchild solution the *b*-boundary contains a singular point every neighborhood of which includes the entire space–time![3] Thus, according to this construction, every event of the space–time—even those in the asymptotic region—is "arbitrarily close" to the singularity. On the other hand, no example has yet been given for which the *g*-boundary seems unphysical. Non-Hausdorff topology from the *g*-boundary construction—and also the *b*-boundary —is known to occur in space–times, such as Taub–NUT, having an incomplete timelike geodesic which continually reenters every small neighborhood of some point of M. However, such a topology seems, physically, to be the "right answer" in such causally pathological space–times, and so they do not provide evidence against these constructions.

Our main purpose is to give an example of a space–time for which the *g*-boundary construction—as well as any construction sufficiently similar to it, including the *b*-boundary—yields an $\bar{M}$ which appears unphysical. Consider any construction of singular points which shares with the *g*-boundary the following two properties: (i) Every incomplete geodesic in the original space–time (M,g_{ab}) terminates at a singular point of $\bar{M}$. (ii) The resulting space $\bar{M}$ is geodesically continuous, in a sense made precise below. We show that such a construction must, when applied to our example, yield the following non-T_1 behavior: There is a singular point s and a regular space–time event r such that every neighborhood of s includes r. In fact, properties significantly weaker than (i) and (ii) suffice. Indeed, as we shall see, unphysical topology results from constructions which merely satisfy these two conditions when applied to all flat space–times.

Geodesic continuity is defined as follows. Recall that the exponential mapping exp takes point p of M together with tangent vector ξ^a at p to that point of M—if it exists—obtained by traversing unit affine length (as determined by ξ^a) along the geodesic with initial conditions (p,ξ^a). Thus, exp is a continuous (in fact, C^∞) mapping from an open subset $\mathcal{O}$ of the tangent bundle of M onto M. This $\mathcal{O}$ consists of (p,ξ^a) for which the corresponding maximally extended geodesic has affine length exceeding one. Let $\bar{\mathcal{O}}$ be the subset which includes in addition those with affine length exactly one. For a construction of singular points satisfying property (i) above, we can extend $\exp: \mathcal{O} \to M$ to a mapping $\overline{\exp}: \bar{\mathcal{O}} \to \bar{M}$ as follows: For $(p,\xi^a) \in \bar{\mathcal{O}} - \mathcal{O}$, let $\overline{\exp}(p,\xi^a)$ be the singular point of $\bar{M}$ representing the endpoint of the incomplete geodesic determined by (p,ξ^a). A construction will be said to be *geodesically continuous* if this extended exponential mapping $\overline{\exp}$ is also continuous. Thus, condition (ii) requires that, for any sequence (p_i,ξ_i^a) in $\bar{\mathcal{O}}$ which converges in $\bar{\mathcal{O}}$ to (p,ξ^a) in $\bar{\mathcal{O}}$, $\overline{\exp}(p_i,\xi_i^a)$ converges in $\bar{M}$ to $\overline{\exp}(p,\xi^a)$. In fact, we shall need this condition only in the case with all the ξ^a's timelike and the sequence (p_i,ξ_i^a) in $\mathcal{O}$ itself.

2. THE EXAMPLE

We first briefly review the *g*-boundary and *b*-boundary constructions and their relevant properties. Let (M,g_{ab}) be a space–time. Each produces from this a topological space $\bar{M}$.

The *g*-boundary[1] construction proceeds as follows. Let

 0022-2488/82/030432-04$02.50 © 1982 American Institute of Physics

$\hat{M}$ denote the disjoint union of M and the set of all incomplete geodesics of unit affine length. Define a mapping $\hat{\exp}:\overline{\mathscr{O}}\to\hat{M}$ in the manner described above for $\overline{\exp}$. Let the topology of $\hat{M}$ be the finest for which $\hat{\exp}$ is continuous, i.e., let the open sets of $\hat{M}$ be those whose inverse images under $\hat{\exp}$ are open in $\overline{\mathscr{O}}$. Next define an equivalence relation, $\sim$, on $\hat{M}$ as follows: Write $x\sim y$, for $x,y\in\hat{M}$, if the neighborhoods of x and y, intersected with M, are identical. Then the g completion of (M,g_{ab}) is $\overline{M}=\hat{M}/\sim$, i.e., the set of equivalence classes under $\sim$, with the topology induced by that of $\hat{M}$. It is manifest from the construction that $\overline{\exp}:\overline{\mathscr{O}}\to\overline{M}$ is continuous. Thus, the g-completion satisfies our two conditions above.

The b-boundary construction proceeds as follows.[2] Denote by B the bundle of orthonormal frames of M, i.e., the ten-dimensional manifold consisting of pairs (p,F) where F is an orthonormal tetrad at p. Introduce on B the following metric tensor field: The squared distance between infinitesimally nearby points (p,F) and (p',F') of B is the sum of the squares of the components, with respect to F, of the connecting vector between p and p', plus the sum of the squares of the F-components of the respective differences of the vectors in F and $\hat{F}'$, where $\hat{F}'$ is the result of parallel transport of F' from p' to p along the connecting vector. This metric tensor is positive-definite, and so B becomes a metric space. The Lorentz group L acts on B, by acting on the frame F keeping the base point fixed (and so $B/L=M$). One checks that this action of L takes Cauchy sequences in B to Cauchy sequences, and so extends to an action on $\overline{B}$, the completion of B as a metric space. The b completion of M is $\overline{M}=\overline{B}/L$, i.e., the set of L-orbits in $\overline{B}$ with the topology induced from that of $\overline{B}$.

This b-completion also satisfies our two conditions. To verify (i), let γ be any geodesic in M of finite affine length. Then the result of parallel transport of a frame F along γ is a curve in B which has finite length there. Thus, this horizontal lift of γ terminates at some point of $\overline{B}$, whence γ terminates at the corresponding point of $\overline{M}$. For condition (ii), let (p_i,ξ_i^a) converge to (p,ξ^a) in $\overline{\mathscr{O}}$. Choose a frame F at p and frames F_i at the p_i, such that the F_i converge to F. Parallel transport each of these frames along its geodesic, to obtain, as above, points x_i and x of $\overline{B}$. But, by smoothness of the exponential map, we have that, for any $K<1$, the points $\exp(p_i,K\xi_i^a)$ of M converge to $\exp(p,K\xi^a)$, as do the corresponding frames parallel-transported to these points. Choosing K near one, it follows that the x_i converge to x in $\overline{B}$. So, projecting to $\overline{M}$, we see that geodesic continuity holds.

We turn now to our example of a causally well-behaved space–time for which every $\overline{M}$ satisfying conditions (i) and (ii) above must have a regular space–time event included in every neighborhood of a singular point. The idea is to introduce a singularity whose character is such that geodesics passing very near it suffer a large "blueshift" and thereby end up far from the singularity.

The example will make use of the observation that any timelike curve can, via a conformal factor, be made a geodesic.

Lemma: Let (M,g_{ab}) be a strongly causal space–time, γ a timelike curve in this space–time, and U an open neighborhood of the portion of γ on which its acceleration is nonzero. Then there exists a smooth positive function Ω with $\Omega=1$ on γ and $\Omega=1$ outside U, such that γ is a geodesic with respect to the metric $\Omega^2 g_{ab}$.

Proof: For τ^a the unit tangent to γ in g_{ab}, and ∇_a and $\tilde{\nabla}_a$ the derivative operations of g_{ab} and $\Omega^2 g_{ab}$, respectively, we have

$$\tau^m\tilde{\nabla}_m\tau^a=\tau^m\nabla_m\tau^a+\Omega^{-1}g^{am}\nabla_m\Omega+2\Omega^{-1}\tau^a\tau^m\nabla_m\Omega.$$

Since $\tau^m\nabla_m\tau^a$ is orthogonal to τ^a, we can consistently set $\Omega=1$ and $\nabla_a\Omega=-g_{ab}\tau^m\nabla_m\tau^b$ on γ, thereby making γ a geodesic in the metric $\Omega^2 g_{ab}$. Strong causality prevents the possible occurrence of any global problems in the definition of Ω, as could arise if γ continually returned arbitrarily close to some space–time point.

Our example is the following. In two-dimensional Minkowski space–time $(N=R^2,\eta_{ab})$, let $s\in N$ and let r lie on a future-directed null geodesic from s. Let γ be a future-directed timelike geodesic of unit length which begins at point p with unit tangent ξ^a, and terminates at s. Let γ_i be a sequence of smooth timelike curves from p to r as shown in Fig. 1. Each is a geodesic from p to a point near s, there has a smooth rapid bend, and then continues as a geodesic to point r. Let the U_i be small nonoverlapping world tubes, each surrounding that portion of its γ_i at which the bend occurs. By the Lemma, choose positive conformal factor Ω with $\Omega=1$ on each γ_i and $\Omega=1$ outside the U_i, and which makes each γ_i a timelike geodesic. Choose Ω smooth everywhere except s (where it cannot be chosen smooth, since the U_i neighborhoods accumulate there). Our four-dimensional example is

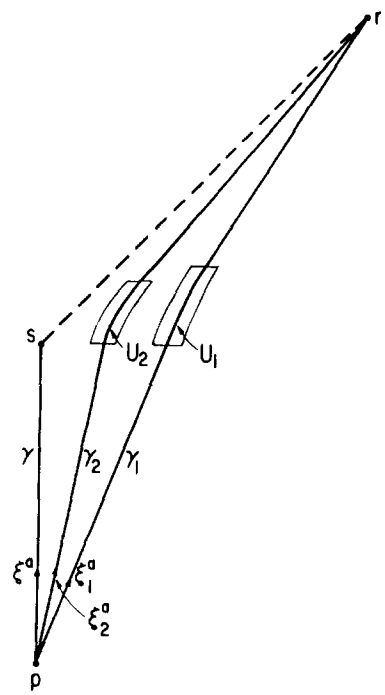

FIG. 1. Minkowski two-space, $N=R^2,\eta_{ab}$. The timelike curves γ_i approach the broken geodesic consisting of γ and the null segment from s to r. Removing point s and applying an appropriate conformal factor equal to one outside of the U_i, there results a space–time for which a large class of constructions of a singular boundary yields a pathological topology.

the cross product of $(N - s, \Omega^2 \eta_{ab})$ and the flat spacelike plane.

Denote by ξ_i^a the tangent to γ_i at p, normalized so that r lies at unit affine length from p along γ_i. Since $\Omega = 1$ on γ_i, the length of γ_i in this space–time is identical to its length in (N, η_{ab}). But in the Minkowski space–time these lengths of the γ_i approach the length of γ as $i \to \infty$. So, we have $\xi_i^a \to \xi^a$ at p. Furthermore, $\exp(p, \xi_i^a) = r$ for all i. Thus, for any geodesically continuous $\bar{M}$, the constant sequence $\{r\}$ must converge to the singular point s associated with the incomplete geodesic γ. In other words, every neighborhood of the singular point s contains the regular space–time event r. That is, $\bar{M}$ fails to be T_1.

In fact, this example can be modified to be flat everywhere. First choose each world tube U_i to be sufficiently narrow that any null geodesic beginning at any point of γ_i at which its acceleration was nonzero meets the boundary of U_i on one of its timelike sides, rather than on one of its spacelike ends. Next, remove from the space–time the timelike sides of the U_i. Then Ω can be chosen, consistent with our earlier conditions, such that $\ln \Omega$ satisfies the η-wave equation everywhere. This follows from the fact that each γ_i is, with the roles of space and time reversed for N, a spacelike surface with U_i in its domain of dependence, while the value of Ω on γ_i (one) together with the value of its normal derivative (that required by the Lemma) is Cauchy data for the wave equation. The resulting solution will be discontinuous across the sides of the U_i, which is the reason these were removed. However, since the U_i are sufficiently narrow the solution will have $\Omega = 1$ on the ends, and so will join smoothly with $\Omega = 1$ outside the U_i. This modification indeed yields a flat space–time, since the wave equation on $\ln \Omega$ ensures in two dimensions the vanishing of the scalar curvature, and hence the flatness, of $\Omega^2 \eta_{ab}$.

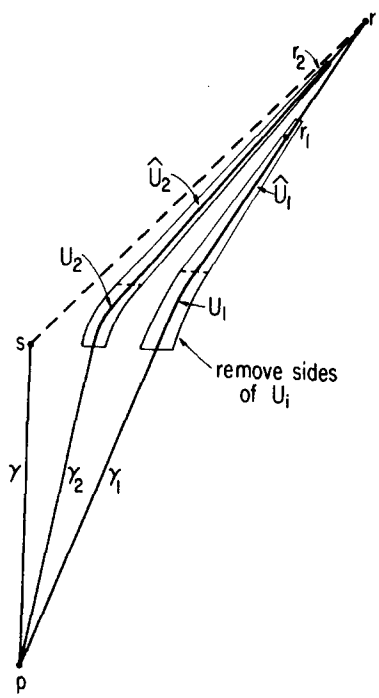

FIG. 2. The space–time of Fig. 1, but with a specific choice of conformal factor, and with the timelike sides of the world tubes U_i removed. There results a flat example having the same features as that of Fig. 1.

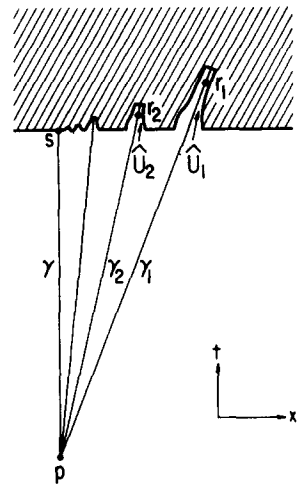

FIG. 3. That portion of Fig. 2 consisting of the lower half-plane together with the $\hat{U}_i$, embedded as the region shown of flat Minkowski two-space. Thus, Fig. 2 is an extension of the space–time shown here. Any construction of a singular boundary for which the r_i converge to s in every extension of Fig. 3 must yield a pathological topology.

One may better visualize the structure of this flat example as follows. Choose points r_i, converging to r, on the γ_i. Let, for each i, $\hat{U}_i$ be an extension of the world tube U_i into the future along γ_i, to include r_i but not r. The $\hat{U}_i$ are to become narrower into the future such that again they do not overlap, as in Fig. 2. The region of this example consisting of the lower half-plane together with the $\hat{U}_i$ is of course a flat, two-dimensional space–time. Furthermore, it is one which can be represented as a region of the Minkowski two-plane. There results Fig. 3, in which the r_i "appear to converge" to the singular point s. So, Fig. 3 is a region of two-dimensional Minkowski space–time, while Fig. 2 is an extension of it. Thus, we may draw the following further conclusion: Any construction of singular points which, applied to Fig. 3, results in the space–time points r_i converging to the singular point s, the terminus of the incomplete geodesic γ—and which maintains this property in any extension—must manifest non-Hausdorff behavior.

3. DISCUSSION

The examples above show that the requirement that geodesic continuity hold—or even that a slightly weaker version of it hold just for flat examples—results in a completed space $\bar{M}$ such that some interior point of the original space–time M lies "arbitrarily close" to the singularity. While the examples are somewhat artificial, it does seem plausible that any construction of singular points which yields such unphysical behavior in these examples will not be generally applicable and useful in an analysis of the structure of singular space–times.

One might hope to find some new construction which avoids the condition of geodesic continuity. A natural candi-

date would be to take the $\bar{M}$ which results from applying the g-boundary or b-boundary construction, and then include as additional open sets in $\bar{M}$ all complements of compact sets in M. This step would automatically eliminate the possibility of an interior space–time point lying in every neighborhood of a singular point. But we may modify our example by removing r, thus making it a new singular point. Then the singular points s and r will be non-T_1 related, which also seems unphysical, though perhaps to a lesser extent. A similar candidate would be to apply the g-boundary or b-boundary construction to every space–time which arises from the given one by removing from it a compact set. Then combine in some way the resulting singular boundaries to obtain $\bar{M}$, retaining in $\bar{M}$ only those singular points which persist, in an appropriate sense, for all choices of the compact set. But we may also modify our example by removing the null line segment from s to r. There would presumably again result singular points s and r, neither identified nor T_1-related. Thus, for both of these candidates, further steps to modify the topology would seem to be appropriate.

Thus, while it seems possible that one could invent mathematically some construction of a singular boundary not subject to obvious physical objections, it is by no means clear that there exists any construction which is both natural and useful in physical problems. The purpose of such a construction, after all, is merely to clarify the discussion of various physical issues involving singular space–times: General relativity as it stands is fully viable with no precise notion of "singular points." Perhaps the localization of singular behavior will go the way of "simultaneity" and "gravitational force."

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation.

[1]R. Geroch, J. Math. Phys. **9**, 450 (1968).
[2]B. G. Schmidt, Gen. Relativ. Gravit. **1**, 269 (1971).
[3]R. A. Johnson, J. Math. Phys. **18**, 898 (1977).

The Riemann tensor, the metric tensor, and curvature collineations in general relativity

C. B. G. McIntosh
Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia

W. D. Halford
Department of Mathematics and Statistics, Massey University, Palmerston North, New Zealand

(Received 15 July 1980; accepted for publication 5 December 1980)

The equation $x_{\mu\nu}R^{\mu}{}_{\lambda\alpha\beta} + x_{\mu\lambda}R^{\mu}{}_{\nu\alpha\beta} = 0$, where $x_{\mu\nu}$ and $R^{\mu}{}_{\nu\alpha\beta}$ are the components of an arbitrary symmetric tensor and of the Riemann tensor formed from the metric tensor $g_{\mu\nu}$, is trivially satisfied by $x_{\mu\nu} = \phi g_{\mu\nu}$. Nontrivial solutions are important in various areas of general relativity such as in the study of curvature collineations, and also in the study of algebraic methods given by Hlavatý and Ihrig for the determination of $g_{\mu\nu}$, from a given set of $R^{\mu}{}_{\nu\alpha\beta}$. We have found all $R^{\mu}{}_{\nu\alpha\beta}$ for which there exist nontrivial solutions of the above equation, and we have given the form of the $x_{\mu\nu}$ in each case. Various examples of space–times for explicit nontrivial solutions are discussed.

PACS numbers: 04.20.Jb, 04.20.Me, 02.40.Ky

1. INTRODUCTION

The equation

$$g_{\mu\nu}R^{\mu}{}_{\lambda\alpha\beta} + g_{\mu\lambda}R^{\mu}{}_{\nu\alpha\beta} = 0, \tag{1.1}$$

where the $g_{\mu\nu}$ are the components of a metric tensor **g** in some local coordinate system and the $R^{\mu}{}_{\nu\alpha\beta}$ are the components of the Riemann tensor **R** formed from the $g_{\mu\nu}$, is of course an identity. It states that the $R_{\mu\nu\alpha\beta}$ are skew-symmetric in the first two indices.

The equation

$$x_{\mu\nu}R^{\mu}{}_{\lambda\alpha\beta} + x_{\mu\lambda}R^{\mu}{}_{\nu\alpha\beta} = 0 \tag{1.2}$$

then has a solution $x_{\mu\nu} = x_{\nu\mu}$ given by

$$x_{\mu\nu} = \phi g_{\mu\nu} \quad (\phi \text{ a scalar}). \tag{1.3}$$

This is the general solution for most sets of $R^{\mu}{}_{\nu\alpha\beta}$ which are formed from a given set of $g_{\mu\nu}$. However, there exist some metrics for which (1.3) is not the general solution of (1.2). In this paper those sets of $R^{\mu}{}_{\nu\alpha\beta}$ are found which are the components of the Riemann tensor **R** for some space–time with metric **g** and for which the general solution of (1.2) is not simply (1.3).

Equation (1.2) arises in general ralativity in different ways. The motivation for our study of (1.2) here is discussed briefly in Secs. 2 and 3 (a fuller discussion is presented elsewhere; see Halford, McIntosh, and van Leeuwen,[1] and McIntosh[2]). In Sec. 4 we describe the problem in detail and in Sec. 5 we outline the formalism which is used to solve our problem. Section 6 together with an Appendix contains the set of solutions, while Sec. 7 analyzes some examples from this list of solutions. Finally, in Sec. 8, we discuss briefly some matters which will be given further investigation.

2. CURVATURE COLLINEATIONS

One place where (1.2) arises is in the study of curvature collineations. A *curvature collineation* (CC) was defined by Katzin, Levine, and Davis[3] to be a vector field ξ for which

$$\mathcal{L}_{\xi}R^{\mu}{}_{\nu\alpha\beta} = 0. \tag{2.1}$$

If the Lie derivative of (1.1) is taken along ξ, then (2.1) ensures that $x_{\mu\nu}$ satisfies (1.2), where

$$\mathcal{L}_{\xi}g_{\mu\nu} = 2\xi_{(\mu;\nu)} = x_{\mu\nu}. \tag{2.2}$$

Thus equation (1.2) is a necessary, but not sufficient condition for ξ to be a CC of **g**. Collinson[4] used this fact to help show that the Weyl tensor of a vacuum space–time metric must be of Petrov type N for the space–time to admit a nontrivial CC. (A *nontrivial* CC is one which does not degenerate to a conformal motion or a homothetic motion or a Killing motion—see Fig. 1 of the paper by Katzin and Levine.[5]) Collinson[4] also showed that for a type N vacuum metric the solution of (1.2) is

$$x_{\mu\nu} = \phi g_{\mu\nu} + \alpha l_{\mu}l_{\nu}, \tag{2.3}$$

where α is an arbitrary scalar and **l** is the repeated principal null vector of the Weyl tensor such that

$$l_{\mu}R^{\mu}{}_{\nu\alpha\beta} = 0. \tag{2.4}$$

Aichelburg[6] and Halford, McIntosh, and van Leeuwen[1] have further examined some type N vacuum metrics which admit nontrivial CC's. Tariq and Tupper[7] showed that the only source-free Einstein–Maxwell space–times which admit nontrivial CC's have a Petrov type N or O Weyl tensor and a null electromagnetic field. Katzin, Levine, and Davis[3] proved that every CC in an Einstein space ($R_{\mu\nu} = \frac{1}{4}Rg_{\mu\nu} \neq 0$) is a Killing motion (isometry), i.e., Einstein spaces do not admit nontrivial CC's.

These results suggest that CC's are usually conformal motions (including the more degenerate subcases), i.e., CC's are usually trivial. But what does "usually" mean? Since for a CC $x_{\mu\nu}$ from (2.2) must satisfy (1.2), only those space–times for which (1.2) has a solution more general than (1.3) can possibly admit nontrivial CC's. It is shown here that those cases are rare, and this result is discussed more fully by one of us elsewhere.[2]

 0022-2488/82/030436-06$02.50 © 1982 American Institute of Physics

3. THE ALGEBRAIC EQUATION

Equation (1.2) is very interesting when looked at simply as an algebraic relationship. If the $R^{\mu}{}_{\nu\alpha\beta}$ are the given components of the Riemann tensor **R** for some metric tensor with components $g_{\mu\nu}$ at a certain point in the manifold, then by using purely algebraic methods Hlavatý[8-10] and Ihrig[11] were able to show that the solution of equation (1.2) at that point is of the form

$$x_{\mu\nu} = \phi g_{\mu\nu} \quad (\phi \text{ a scalar}) \tag{3.1}$$

except where, in Ihrig's language, **R** is not total at that point. (The Riemann tensor **R** is said to be total at a point if the dimension of the space of bivectors which span the curvature 2-form is equal to the dimension of the Lorentz group at that point.) When **R** is total at a point, Ihrig's method, which uses only algebra, enables one to construct the components $g_{\mu\nu}$ at that point from the given $R^{\mu}{}_{\nu\alpha\beta}$ up to a conformal factor. His method can be extended to determine those metrics whose Riemann tensor is not total; this will be discussed elsewhere in relation to the holonomy group of the space–time manifold. Hlavatý used the connection between the holonomy group and the integrability conditions of $g_{\mu\nu;\lambda} = 0$ in his treatment of Eqs. (1.1) and (1.2).

4. THE PROBLEM

The problem then is how to find all sets of $R^{\mu}{}_{\nu\alpha\beta}$ formed from some space–time metric $g_{\mu\nu}$ for which

$$x_{\mu\nu}R^{\mu}{}_{\lambda\alpha\beta} + x_{\mu\lambda}R^{\mu}{}_{\nu\alpha\beta} = 0 \quad (x_{\mu\nu} = x_{\nu\mu}), \tag{4.1}$$

does not have just the trivial solution $x_{\mu\nu} = \phi g_{\mu\nu}$, and to determine the forms of $x_{\mu\nu}$ in those cases. This problem is equivalent to the following: *Given all possible forms of* $x_{\mu\nu}$, *which ones have sets of* $R^{\mu}{}_{\nu\alpha\beta}$ *associated with them such that (4.1) is satisfied and for which the* $x_{\mu\nu}$ *are not just trivial solutions*? From this point of view **x** is an arbitrary symmetric second order tensor in space–time, and so can be classified in some standard way as can every such tensor. There are several classification schemes available. We shall use the one developed by Plebański.[12] The formalism employed in this paper is the null tetrad formalism of Newman and Penrose (NP).[13] The detailed classification of a tensor such as **x** in NP language according to Plebański's scheme is given by McIntosh, Foyster, and Lun,[14] and is used here.

Equation (4.1) can be written in the tetrad formalism with basis 1-forms θ^a, described in the next section, as

$$x_{ab}\Theta^{a}{}_{c} + x_{ac}\Theta^{a}{}_{b} = 0, \tag{4.2}$$

where the curvature 2-forms $\Theta^{a}{}_{b}$ are defined in terms of the tetrad components $R^{a}{}_{bcd}$ of the Riemann tensor by

$$\Theta^{a}{}_{b} = \tfrac{1}{2}R^{a}{}_{bcd}\theta^{c}\wedge\theta^{d}. \tag{4.3}$$

All small Roman indices take values 1, 2, 3, 4. The procedure is to take a particular Plebański class for **x**, write x_{ab} in a canonical form for that class, and find the nonzero components $\Theta^{a}{}_{b}$ (and hence $R^{a}{}_{bcd}$) for that x_{ab}. The various Plebański classes will be examined one by one in this way. The nonzero components $R^{a}{}_{bcd}$ for each class will be given in terms of the NP tetrad components Ψ_A ($A = 1,2,3,4$), Φ_{AB} ($A,B = 0,1,2$), and Λ. The Ψ_A and Φ_{AB} are respectively the components of the Weyl tensor and the trace-free Ricci tensor, while $\Lambda = R/24$ where R is the scalar curvature.

5. FORMALISM AND BASIC EQUATIONS

The space–time manifold is spanned by four null vectors **l**, **n**, **m**, and $\bar{\mathbf{m}}$, where **l** and **n** are real and **m** is complex. We take a set of basis 1-forms θ^a to be

$$\theta^1 = \mathbf{n}, \quad \theta^2 = \mathbf{l}, \quad \theta^3 = -\bar{\mathbf{m}}, \quad \theta^4 = \bar{\theta}^3 = -\mathbf{m}. \tag{5.1}$$

The metric is then

$$ds^2 = 2(\theta^1\theta^2 - \theta^3\theta^4) = g_{ab}\theta^a\theta^b. \tag{5.2}$$

A basis for a space of complex self-dual 2-forms is given by

$$Z^1 = \theta^1\wedge\theta^3, \quad Z^2 = \theta^1\wedge\theta^2 - \theta^3\wedge\theta^4, \quad Z^3 = \theta^4\wedge\theta^2. \tag{5.3}$$

It then follows from Newman and Penrose[13] and from Debever, McLenaghan, and Tariq[15] that the $\Theta^{a}{}_{b}$ defined by (4.3) are given by

$$\Theta^{4}{}_{1} = \Theta^{2}{}_{3} = \Psi_0 Z^1 + \Psi_1 Z^2 + (\Psi_2 + 2\Lambda)Z^3 + \Phi_{00}\bar{Z}^1 + \Phi_{01}\bar{Z}^2 + \Phi_{02}\bar{Z}^3, \tag{5.4a}$$

$$\Theta^{2}{}_{2} = -\Theta^{1}{}_{1} = (\Psi_1 + \Phi_{01})Z^1 + (\Psi_2 - \Lambda + \Phi_{11})Z^2 + (\Psi_3 + \Phi_{21})Z^3 + (\bar{\Psi}_1 + \Phi_{10})\bar{Z}^1 + (\bar{\Psi}_2 - \Lambda + \Phi_{11})\bar{Z}^2 + (\bar{\Psi}_3 + \Phi_{12})\bar{Z}^3, \tag{5.4b}$$

$$\Theta^{4}{}_{4} = -\Theta^{3}{}_{3} = (\Psi_1 - \Phi_{01})Z^1 + (\Psi_2 - \Lambda - \Phi_{11})Z^2 + (\Psi_3 - \Phi_{21})Z^3 - (\bar{\Psi}_1 - \Phi_{10})\bar{Z}^1 - (\bar{\Psi}_2 - \Lambda - \Phi_{11})\bar{Z}^2 - (\bar{\Psi}_3 - \Phi_{12})\bar{Z}^3, \tag{5.4c}$$

$$\Theta^{1}{}_{4} = \Theta^{3}{}_{2} = -(\Psi_2 + 2\Lambda)Z^1 - \Psi_3 Z^2 - \Psi_4 Z^3 - \Phi_{20}\bar{Z}^1 - \Phi_{21}\bar{Z}^2 - \Phi_{22}\bar{Z}^3. \tag{5.4d}$$

The other nonzero $\Theta^{a}{}_{b}$ can be found by taking the complex conjugates of (5.4a) and (5.4d).

The tetrad components of **x** are as follows:

$$x_{11} = x_{\mu\nu}l^{\mu}l^{\nu}, \quad x_{13} = x_{\mu\nu}l^{\mu}m^{\nu},$$
$$x_{1234} \equiv x_{12} + x_{34} = x_{\mu\nu}(l^{\mu}n^{\nu} + m^{\mu}\bar{m}^{\nu}), \tag{5.5}$$

together with corresponding expressions for $x_{14}, x_{22}, x_{23}, x_{24}$, x_{33}, and x_{44}. The term $x_{12} - x_{34}$ given by

$$2(x_{12} - x_{34}) = x_{\mu\nu}g^{\mu\nu} \equiv x \tag{5.6}$$

does not appear in Eqs. (4.2) when written out explicitly since $g_{\mu\nu}$ trivially satisfies the identity (1.1). Equations (4.2) are then 54 real equations for the nine unknowns in x_{ab}, i.e., six equations from the coefficients of Z^i, $\bar{Z}^i$ ($i = 1,2,3$) for each pair (b,c), with (b,c) = (1,2) being equivalent to (b,c) = (3,4). These equations are written out in the Appendix as 27 complex equations.

6. SOLUTIONS FOR x

The canonical forms of the x_{ab} for the various Plebański classes can now be taken one by one and inserted into the 27 Eqs. (A1)–(A6) of the Appendix. For each class the possible

nonzero Ψ_A, Φ_{AB}, and Λ together with any relationship between them are readily obtained and are listed in Table I.

For a second-order symmetric tensor S a fourth-order tensor $\mathscr{P}$ can be constructed from quadratic products of the trace-free tensor X formed from S. Let

$$X_{\mu\nu} = S_{\mu\nu} - \tfrac{1}{4}g_{\mu\nu}S, \tag{6.1}$$

where $S = S_{\mu\nu}g^{\mu\nu}$. Then form the trace-free $\mathscr{P}$ with components

$$\mathscr{P}^{\alpha\beta}{}_{\gamma\sigma} = X^{[\alpha}{}_{[\gamma}X^{\beta]}{}_{\sigma]} + \delta^{[\alpha}{}_{[\gamma}X_{\sigma]\lambda}X^{\beta]\lambda} - \tfrac{1}{6}\delta^{[\alpha}{}_{[\gamma}\delta^{\beta]}{}_{\sigma]}X_{\mu\nu}X^{\mu\nu}. \tag{6.2}$$

Such a tensor was used by Plebański[12] in his classification of the Ricci tensor. Now $\mathscr{P}$ has the same symmetries as the Weyl tensor, and therefore the same Petrov classification scheme can be applied to it. It is discussed by McIntosh, Foyster, and Lun[14] and is there called the *Plebański tensor*. By choosing S in (6.1) to be (a) the tensor $x_{\mu\nu}$, (b) the Ricci tensor $R_{\mu\nu}$, we can form via (6.2) two Plebański tensors in terms of (a) the x_{ab}, (b) the Φ_{AB} respectively.

Let $\mathscr{W}$ denote the Weyl tensor formed from the nonzero Ψ_A. Let $\mathscr{P}_x$ and $\mathscr{P}_\Phi$ denote the Plebański tensors formed from the x_{ab} and from the Φ_{AB} respectively. Table I lists each of the Petrov types of $\mathscr{W}$, $\mathscr{P}_x$, and $\mathscr{P}_\Phi$ for each Plebański class; it is clear that there is a close link between them. The dimension of the space of bivectors which span the curvature 2-forms $\Theta^a{}_b$ is also given in Table I for each Plebański class of the $\mathscr{P}_x$. Subscases are possible when some of the Ψ_A, Φ_{AB}, and Λ listed are zero, but the maximum dimension is given.

The following results emerge from Table I:

(i) There are no nonflat space–times with nontrivial solutions x_{ab} of (4.2) with $\mathscr{P}_x$ of Petrov type I, II, or III.

(ii) The only space–times with nontrivial solutions x_{ab} of (4.2) with $\mathscr{P}_x$ of Petrov type D have both $\mathscr{P}_\Phi$ and $\mathscr{W}$ of Petrov type D.

(iii) The only space–times with nontrivial solutions x_{ab} of (4.2) with $\mathscr{P}_x$ of Petrov type N have $\mathscr{W}$ and $\mathscr{P}_\Phi$ of Petrov type N and O respectively.

(iv) The space–times with nontrivial solutions x_{ab} of (4.2) with $\mathscr{W}$ of Petrov type I or II have $\mathscr{P}_\Phi$ of the same Petrov type. In these cases $\mathscr{P}_x$ is of Petrov type O.

(v) In all the cases (ii)–(iv) where the Petrov types of at least two of $\mathscr{W}$, $\mathscr{P}_x$, and $\mathscr{P}_\Phi$ are the same and are algebraically special, the repeated principal null directions of these Weyl and/or Plebański tensors align.

(vi)The maximum dimension of the space of bivectors which span the curvature 2-forms $\Theta^a{}_b$ in any one Plebański class is three. This can be seen by taking the nonzero Ψ_A, Φ_{AB}, and Λ and the relationships between them in each class and constructing the $\Theta^a{}_b$ from (5.4). For instance, when x is in the class $[3T\text{–}S]_2$ and $\Theta^a{}_b$ are spanned by $Z^1 + \bar{Z}^1$, $Z^2 + \bar{Z}^2$, and $Z^3 + \bar{Z}^3$.

The last point is perhaps the most interesting and leads to the theorem of McIntosh[2] that nontrivial CC's can only occur in space–times whose curvature 2-forms are spanned by at most three linearly independent bivectors. It was known from the work of Ihrig[11] that the maximum dimension of this bivector space was five if the space–time was such

TABLE I. Solutions for x according to Plebański class.

Plebański class of x	Petrov type of $\mathscr{P}_x$	Nonzero x_{ab}	Nonzero $\Psi_A, \Phi_{AB}, \Lambda$	Dimension bivector space for $\Theta^a{}_b$	Petrov type of $\mathscr{W}$	Petrov type of $\mathscr{P}_\Phi$
$[T\text{-}S_1\text{-}S_2\text{-}S_3]_4$	I	$x_{11}=x_{22}, x_{1234}, x_{33}=x_{44}$,	—		O	O
$[2T\text{-}S_1\text{-}S_2]_3$	D	$x_{1234}, x_{33}=x_{44}$	$\Psi_2=-2\Lambda,\ \Phi_{11}=-3\Lambda$	1	D	D
$[T\text{-}2S_1\text{-}S_2]_3$	D	$x_{1234}, x_{11}=x_{22}$	$\Psi_2=-2\Lambda,\ \Phi_{11}=3\Lambda$	1	D	D
$[2T\text{-}2S]_2$	D	x_{1234}	$\Psi_2=-2\Lambda,\ \Phi_{11}$	2	D	D
$[3T\text{-}S]_2$	O	$x_{33}=x_{44}=-x_{1234}$ [or $x_{11}=x_{22}=-x_{1234}$]	$\Psi_0=\Phi_{00}, \Psi_4=\Phi_{22}, \Psi_2+2\Lambda=\Phi_{02}=\Phi_{20}$, $\Psi_2=\Phi_{11}+\Lambda, \Psi_1=\bar{\Psi}_1=\Phi_{01}, \Psi_3=\bar{\Psi}_3=\Phi_{21}$	3	I	I
$[T\text{-}3S]_2$	O	$x_{11}=x_{22}=x_{1234}$	$\Psi_0=\bar{\Psi}_4=\Phi_{02}, \Phi_{00}=\Phi_{22}=\Psi_2+2\Lambda$, $\Psi_2=\Lambda-\Phi_{11}, \Psi_1=-\bar{\Psi}_3=-\Phi_{01}=\Phi_{12}$	3	I	I
$[4T]_1$	O	—	any	6	I	I
$[Z\text{-}\bar{Z}\text{-}S_1\text{-}S_2]_4$	I	$x_{11}=-x_{22}, x_{1234}, x_{33}=x_{44}$,	—		O	O
$[Z\text{-}\bar{Z}\text{-}2S]_3$	D	$x_{11}=-x_{22}, x_{1234}$	$\Psi_2=-2\Lambda,\ \Phi_{11}=3\Lambda$	1	D	D
$[2N\text{-}S_1\text{-}S_2]_4$	II	$x_{22}, x_{1234}, x_{33}=x_{44}$	—		O	O
$[2N\text{-}2S]_{[2\text{-}1]}$	D	x_{22}, x_{1234}	$\Psi_2=-2\Lambda, \Phi_{11}=3\Lambda$	1	D	D
$[3N\text{-}S]_3$	N	$x_{22}, x_{33}=x_{44}=x_{1234}$	$\Psi_4=\Phi_{22}$	1	N	O
$[4N]_2$	O	x_{22}	$\Psi_4, \Psi_3=-\Phi_{21}, \Psi_2=-2\Lambda, \Phi_{11}=3\Lambda, \Phi_{22}$	3	II	II
$[3N\text{-}S]_4$	III	$x_{33}=x_{44}=-x_{1234}$, $x_{23}\neq -x_{24}$	—		O	O
$[4N]_3$	N	$x_{23}=-x_{24}$	$\Psi_4=\Phi_{22}$	1	N	O

that (4.1) has a nontrivial solution for **x**. However, the calculation of all possible sets of $\mathcal{O}^a{}_b$ and x_{ab} for (4.2), equivalently (4.1), which have nontrivial solutions has reduced this maximum dimension to three. There is no reason, other than this computational one, yet known why this should be so.

Other interesting points include the lack of nontrivial solutions for **x** when the Petrov type of $\mathcal{W}$, $\mathcal{P}_x$, and $\mathcal{P}_\Phi$ simultaneously is I or II or III, and the alignment of the repeated principal null directions of these tensors causing the groupings D-D-D or N-N-O to appear. It is not yet fully known why this occurs—again these results just appear out of the calculations which led to Table I.

The type of alignment of repeated null directions mentioned in the last paragraph occurs elsewhere in general relativity in a completely different circumstance. MacCallum[16] discusses locally isotropic space–times with non-null homogeneous hypersurfaces. In all the space–times which arise in this work the Petrov types of the Weyl tensor $\mathcal{W}$ and the Plebański tensor $\mathcal{P}_\Phi$ are the same, and also the repeated principal null directions of these tensors align. (There is a brief comment on this in the paper by McIntosh, Foyster, and Lun[14]; it will be discussed more fully elsewhere.)

7. EXAMPLES

No attempt is made here to determine all those space-times whose tetrad components of their Riemann tensors are listed in Table I. However, it is interesting to look at some space–times with sets of $R^\mu{}_{\nu\alpha\beta}$ such that (1.2) has nontrivial solutions for $x_{\mu\nu}$. These are as follows:

(i) *Vacuum metrics.* As mentioned in Sec. 2 Collinson[4] showed that the only vacuum space–times with nontrivial solutions to (1.2) are Petrov type N ones with the $x_{\mu\nu}$ given by (2.3) and (2.4). This result follows immediately from Table I. The Plebański class of the trace-free **x** is $[4N]_2$ and the nonzero tetrad component is x_{22}. This is readily shown to be equivalent to (2.3). In the case of the well-known plane-fronted waves of this type Katzin, Levine, and Davis[17] and Aichelburg[6] showed that for the nonrotating or pp-wave metrics there always exists a vector field ξ which satisfied (2.1) and (2.2) and hence is a CC of these metrics. Halford, McIntosh, and van Leeuwen[1] extended this result to the rotating case given by Kundt[18] and gave the explicit form of the CC in this case.

(ii) *Einstein–Maxwell solutions.* Table I agrees with Tariq and Tupper's result[7] and nontrivial CC's, equivalent here to nontrivial solutions of (1.2), can be found for source-free Einstein–Maxwell space–times only when the Maxwell field is null and the Weyl tensor is Petrov type N or O. In tetrad form these are the cases with $\Phi_{22}\neq 0$, $\Psi_4\neq 0$ or $\Phi_{22}\neq 0$, $\Psi_4 = 0$ respectively. Both cases admit x_{22} only. There is an extra case with $\Psi_4 = \Phi_{22}$ in which **x** could be in the Plebański class $[3N\text{-}S]_3$ or $[4N]_3$. There is a mistake in Tariq and Tupper's paper[7] in that their equations under the heading of null electromagnetic fields do not agree with equations (A1–A6) in the Appendix of our present paper. However, their result is unaffected by the correction of the appropriate equations.

(iii) *Conformally flat space–times.* If all the Ψ_A are zero and the space–time is conformally flat, then (1.2) can have various types of solution with the nonzero Φ_{AB} and Λ satisfying one of the cases

$$\begin{aligned}&\text{(a) } \Phi_{11} \text{ arbitrary,} \qquad \text{(b) } \Phi_{22} \text{ arbitrary,}\\ &\text{(c) } \Phi_{11} = \pm 3\Lambda, \quad \text{(d) } \Phi_{11} = \Lambda,\ \Phi_{00} = \Phi_{22} = \pm 2\Lambda.\end{aligned} \tag{7.1}$$

Katzin, Levine, and Davis[19] discussed CC's in conformally flat space–times and showed that

$$x_{\mu\nu} = \phi g_{\mu\nu} + \rho R_{\mu\nu}, \tag{7.2}$$

where ρ is an arbitrary scalar and $R_{\mu\nu}$ satisfies several side conditions. Now the conditions (7.1) are each equivalent to (7.2), and the side conditions on the $R_{\mu\nu}$ arise when (2.1) and (2.2) are to be satisfied as well for a CC to occur.

(iv) *Einstein space–times.* These are characterized by $R_{\mu\nu} = 6\Lambda g_{\mu\nu}$, $\Lambda \neq 0$. In tetrad form this is equivalent to $\Phi_{AB} = 0$, $\Lambda \neq 0$ for all A, B. Table I shows clearly that there are no nontrivial solutions **x** of (1.2) in this case, i.e., nontrivial CC's do not exist in Einstein space–times. This confirms the results of Katzin, Levine, and Davis[3] and of Tariq and Tupper.[7]

(v) *Gödel metric.* The Gödel[20] cosmological metric is

$$ds^2 = (dt + e^{ax}dy)^2 - dx^2 - dz^2 - \tfrac{1}{2}e^{2ax}dy^2, \tag{7.3}$$

where a is a constant. A null tetrad can be found such that the nonzero NP components of the Riemann tensor are

$$\Psi_0 = \Psi_4 = 3\Psi_2 = \Phi_{00} = \Phi_{22} = 2\Phi_{11} = -6\Lambda = a^2/4.$$

Thus $\Psi_2 + 2\Lambda = 0$ and $\Psi_2 = \Phi_{11} + \Lambda$. In this case **x** is in the Plebański class $[3T\text{-}S]_2$.

(vi) *Metrics of Bertotti–Robinson type.* Consider the metric

$$ds^2 = 2f^2 du dv - 2g^2 d\zeta d\bar{\zeta}, \tag{7.4}$$

where

$$f = (\alpha\zeta + \bar{\alpha}\bar{\zeta} + d)B(u,v), \quad g = (au + bv + c)A(\zeta,\bar{\zeta}), \tag{7.5}$$

with a, b, c arbitrary real constants, α an arbitrary complex constant, and A, B arbitrary functions of their respective arguments. If in (7.5)

$$\begin{aligned}&\alpha = a = b = 0, \quad c = d = 1,\\ &A = (1 + K_1\zeta\bar{\zeta})^{-1}, \quad B = (1 + K_2 uv)^{-1},\end{aligned} \tag{7.6}$$

then the metric becomes one discussed by Plebański[21] [his equations (4.14) and (4.16)]. There exists a tetrad for which

$$\begin{aligned}&\Psi_2 = -2\Lambda, \quad 2\Phi_{11} = K_1 + K_2 = \mathscr{E}^2 + \mathscr{B}^2,\\ &6\Lambda = K_1 - K_2 = \lambda,\end{aligned} \tag{7.7}$$

where $\mathscr{E}$ and $\mathscr{B}$ represent the electromagnetic field and λ is the cosmological constant. Two particular cases of (7.6) and (7.7) are noteworthy. The conformally flat vacuum Einstein–Maxwell space–time found by Bertotti[22] and Robinson[23] is characterized by

$$\Psi_2 = \Lambda = 0, \quad \Phi_{11} = K_1 \quad (K_1 = K_2) \tag{7.8}$$

(See also Tariq and Tupper[24]). Two Petrov type D Einstein space–times are characterized by

$$\Psi_2 + 2\Lambda = 0, \quad \Phi_{11} = 0, \quad \Lambda = K_1/3, \quad (K_2 = -K_1). \tag{7.9}$$

The solutions (7.6) are $V_2 \times V_2$ space–times, each V_2 having constant positive curvature K_1 or K_2 respectively; the Ein-

stein space–times (7.9) have been discussed by Petrov[25] [his Eqs. (14.7) and (14.8)], and (7.8) and (7.9) by Tariq and Tupper[7] (their Sec. 5). Since (7.7) holds, from Table I all these metrics admit a solution **x** of equation (1.2) of the type $[2T\text{-}2S]_2$.

Other special cases of interest occur for (7.4) and (7.5) when

$$\text{(a) } f = 1 \text{ (i.e., } \alpha = 0,\ d = B = 1) \tag{7.10a}$$

or

$$\text{(b) } g = 1 \text{ (i.e., } a = b = 0,\ c = A = 1). \tag{7.10b}$$

There exists a tetrad such that in these cases

$$\text{(a) } \Psi_2 = -2\Lambda,$$
$$\Phi_{11} = 3\Lambda = \tfrac{1}{2}g^{-2}[abA^2 - (\log A)_{,\xi\bar{\xi}}], \tag{7.11a}$$
$$\text{(b) } \Psi_2 = -2\Lambda,$$
$$\Phi_{11} = -3\Lambda = \tfrac{1}{2}f^{-2}[\alpha\bar{\alpha}B^2 - (\log B)_{,u\bar{v}}]. \tag{7.11b}$$

For $\Lambda \neq 0$ the metrics in these two cases obviously have Weyl tensors of Petrov type D. These metrics then have

$$\Psi_2 = 2\Lambda = 0, \quad \Phi_{11} = \pm 3\Lambda.$$

These possibilities occur a number of times in Table I where it is seen that **x** can be of type $[2T\text{-}S_1\text{-}S_2]_3$ or $[3T\text{-}S]_2$ for $\Phi_{11} = -3\Lambda$, and of type $[T\text{-}2S_1\text{-}S_2]_3$, $[3T\text{-}S]_2$, $[T\text{-}3S]_2$, $[Z\text{-}\bar{Z}\text{-}2S]_3$, $[2N\text{-}2S]_{(2\text{-}1)}$, or $[4N]_2$ for $\Phi_{11} = 3\Lambda$. Also, **x** could be of type $[2T\text{-}2S]_2$ or $[4T]_1$ in either case.

(vii) *Space–times admitting a certain type of vector field.* McIntosh and van Leeuwen[26] examine space–times admitting a vector field **v** which is such that

$$v_\mu R^\mu{}_{\nu\alpha\beta} = 0. \tag{7.12}$$

Every space–time of this kind automatically has nontrivial solutions **x** of (1.2) of the form

$$x_{\mu\nu} = \phi g_{\mu\nu} + \alpha v_\mu v_\nu. \tag{7.13}$$

If the space–time is empty, then **v** must be the vector I tangent to the fourfold repeated principal null direction of the Weyl tensor of the metric, as Collinson[4] proved. For other space–times McIntosh and van Leeuwen show that if **v** is null, timelike or spacelike, then **x** is in the Plebański class $[4N]_2$, $[T\text{-}3S]_2$, or $[3S\text{-}T]_2$, respectively.

8. FURTHER COMMENTS

We have seen that if the components of the Riemann tensor of some space–time are known in some coordinate system, then (1.2) almost always determines the metric tensor components $g_{\mu\nu}$ up to a conformal factor. The cases where this does not happen are those space–times which have a Riemann tensor which has tetrad components that satisfy the conditions of one of the cases in Table I (except of course when **x** is in the Plebański class $[4T]_1$, (i.e., $x_{ab} = 0$!).

Ihrig's procedure,[11] which determines the $g_{\mu\nu}$ up to a conformal factor from a given set of $R^\mu{}_{\nu\alpha\beta}$, can be applied in slightly modified form to generate, for example, the pp-wave metric. Relative to a coordinate system $(x^0,x^1,x^2,x^3) = (u,v,\zeta,\bar{\zeta})$, let it be given that

$$R^1{}_{202} = \tfrac{1}{2}F_{,\zeta\zeta}, \quad R^1{}_{303} = \tfrac{1}{2}\bar{F}_{,\bar{\zeta}\bar{\zeta}}. \tag{8.1}$$

together with other nonzero $R^\mu{}_{\nu\alpha\beta}$ obtained from these by using the symmetries of the Riemann tensor. Here $F(u,\zeta)$ is an arbitrary complex function of its arguments. Then using the procedure referred to one obtains

$$g_{00} = G(u,v,\zeta,\bar{\zeta}), \quad g_{01} = \phi, \quad g_{23} = -\phi, \tag{8.2}$$

which gives

$$ds^2 = \phi ds^2_{\text{pp-waves}} + \alpha du^2 = x_{\mu\nu}dx^\mu dx^\nu, \tag{8.3}$$

where

$$ds^2_{\text{pp-waves}} = [F(u,\zeta) + \bar{F}(u,\bar{\zeta})]du^2 + 2dudv - 2d\zeta d\bar{\zeta} \tag{8.4}$$

is the pp-wave metric, α and ϕ are arbitrary functions of the four coordinates, and $G = \alpha + \phi(F + \bar{F})$. This agrees entirely with (2.3), where the $g_{\mu\nu}$ are the pp-wave metric tensor components in the given coordinate system. The generation of (8.2) from (8.1) and similar cases is discussed more fully by McIntosh and Halford.[27]

The link between the Weyl tensor and the Plebański tensor in the examples discussed by MacCallum[16] (see Sec. 6), which all admit an isotropy group, may suggest that the space–times giving nontrivial solutions **x** to equations (1.2) have a large amount of symmetry such as an isotropy group. This is true for some of the examples presented here such as the Gödel metric and the metric (7.9) of Bertotti–Robinson type with cosmological constant, but is not true in general. For instance, the pp-wave metric (8.4) and the metric (7.4), (7.5) with $f = 1$, $abc \neq 0$, and A an arbitrary function of its arguments, admit only one isometry (Killing vector field) each, in general.

APPENDIX

The curvature 2-forms $\Theta^a{}_b$ have been defined in Eq. (4.3) and the trace-free x_{ab} have been given in (5.5) (note that $x_{1234} \equiv x_{12} + x_{34}$). Equations

$$x_{ab}\Theta^a{}_c + x_{ac}\Theta^a{}_b = 0 \tag{4.2}$$

are displayed below as 27 independent complex equations, grouped according to the values of the index pair (b,c):

$(b,c) = (1,1)$:

$$x_{11}(\Psi_2 + \Phi_{11} - \Lambda) - x_{13}\Phi_{10} - x_{14}\Psi_1 = 0, \tag{A1a}$$
$$x_{11}(\Psi_1 + \Phi_{01}) - x_{13}\Phi_{00} - x_{14}\Psi_0 = 0, \tag{A1b}$$
$$x_{11}(\Psi_3 + \Phi_{21}) - x_{13}\Phi_{20} - x_{14}(\Psi_2 + 2\Lambda) = 0. \tag{A1c}$$

$(b,c) = (2,2)$:

$$x_{22}(\Psi_2 + \Phi_{11} - \Lambda) - x_{23}\Psi_3 - x_{24}\Phi_{12} = 0, \tag{A2a}$$
$$x_{22}(\Psi_1 + \Phi_{01}) - x_{23}(\Psi_2 + 2\Lambda) - x_{24}\Phi_{02} = 0, \tag{A2b}$$
$$x_{22}(\Psi_3 + \Phi_{21}) - x_{23}\Psi_4 - x_{24}\Phi_{22} = 0. \tag{A2c}$$

$(b,c) = (1,2)$[or (3,4)]:

$$x_{13}\Psi_3 + x_{14}\Phi_{12} - x_{23}\Phi_{10} - x_{24}\Psi_1 = 0, \tag{A3a}$$
$$x_{13}(\Psi_2 + 2\Lambda) + x_{14}\Phi_{02} - x_{23}\Phi_{00} - x_{24}\Psi_0 = 0, \tag{A3b}$$
$$x_{13}\Psi_4 + x_{14}\Phi_{22} - x_{23}\Phi_{20} - x_{24}(\Psi_2 + 2\Lambda) = 0. \tag{A3c}$$

$(b,c) = (1,3)$:

$$x_{11}\bar{\Psi}_3 - x_{1234}\Phi_{01} + 2x_{13}\Phi_{11} - x_{33}\bar{\Psi}_1 = 0, \quad \text{(A4a)}$$

$$x_{11}\Phi_{12} - x_{1234}\Psi_1 + 2x_{13}(\Psi_2 - \Lambda) - x_{33}\Phi_{10} = 0, \quad \text{(A4b)}$$

$$x_{11}\Phi_{02} - x_{1234}\Psi_0 + 2x_{13}\Psi_1 - x_{33}\Phi_{00} = 0, \quad \text{(A4c)}$$

$$x_{11}(\bar{\Psi}_2 + 2\Lambda) - x_{1234}\Psi_{00} + 2x_{13}\Phi_{10} - x_{33}\bar{\Psi}_0 = 0, \quad \text{(A4d)}$$

$$x_{11}\bar{\Psi}_4 - x_{1234}\Phi_{02} + 2x_{13}\Phi_{12} - x_{33}(\bar{\Psi}_2 + 2\Lambda) = 0, \quad \text{(A4e)}$$

$$x_{11}\Phi_{22} - x_{1234}(\Psi_2 + 2\Lambda) + 2x_{13}\Psi_3 - x_{33}\Phi_{20} = 0. \quad \text{(A4f)}$$

$(b,c) = (2,3)$:

$$x_{22}\Phi_{01} - x_{1234}\bar{\Psi}_3 + 2x_{23}(\bar{\Psi}_2 - \Lambda) - x_{33}\Phi_{21} = 0, \quad \text{(A5a)}$$

$$x_{22}\Psi_1 - x_{1234}\Phi_{12} + 2x_{23}\Phi_{11} - x_{33}\Psi_3 = 0, \quad \text{(A5b)}$$

$$x_{22}\Psi_0 - x_{1234}\Phi_{02} + 2x_{23}\Phi_{01} - x_{33}(\Psi_2 + 2\Lambda) = 0, \quad \text{(A5c)}$$

$$x_{22}\Phi_{02} - x_{1234}\bar{\Psi}_4 + 2x_{23}\bar{\Psi}_3 - x_{33}\Phi_{22} = 0, \quad \text{(A5d)}$$

$$x_{22}\Phi_{00} - x_{1234}(\bar{\Psi}_2 + 2\Lambda) + 2x_{23}\bar{\Psi}_1 - x_{33}\Phi_{20} = 0, \quad \text{(A5e)}$$

$$x_{22}(\Psi_2 + 2\Lambda) - x_{1234}\Phi_{22} + 2x_{23}\Phi_{21} - x_{33}\Psi_4 = 0. \quad \text{(A5f)}$$

$(b,c) = (3,3)$:

$$x_{13}\bar{\Psi}_3 - x_{23}\Phi_{01} + x_{33}(-\bar{\Psi}_2 + \Phi_{11} + \Lambda) = 0, \quad \text{(A6a)}$$

$$x_{13}\Phi_{12} - x_{23}\Psi_1 + x_{33}(\Psi_2 - \Phi_{11} - \Lambda) = 0, \quad \text{(A6b)}$$

$$x_{13}\Phi_{02} - x_{23}\Psi_0 + x_{33}(\Psi_1 - \Phi_{01}) = 0, \quad \text{(A6c)}$$

$$x_{13}(\bar{\Psi}_2 + 2\Lambda) - x_{23}\Phi_{00} + x_{33}(-\bar{\Psi}_1 + \Phi_{10}) = 0, \quad \text{(A6d)}$$

$$x_{13}\bar{\Psi}_4 - x_{23}\Phi_{02} + x_{33}(-\bar{\Psi}_3 + \Phi_{12}) = 0, \quad \text{(A6e)}$$

$$x_{13}\Phi_{22} - x_{23}(\Psi_2 + 2\Lambda) + x_{33}(\Psi_3 - \Phi_{21}) = 0. \quad \text{(A6f)}$$

[1] W. D. Halford, C. B. G. McIntosh, and E. H. van Leeuwen, J. Phys. A: Gen. Phys. **13**, 2995 (1980).
[2] C. B. G. McIntosh, "Curvature collineations in general relativity are almost always special conformal motions," preprint (1980).
[3] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).
[4] C. D. Collinson, J. Math. Phys. **11**, 818 (1970).
[5] G. H. Katzin and J. Levine, Colloq. Math. **26**, 21 (1972).
[6] P. C. Aichelburg, Gen. Relativ. Gravit. **3**, 397 (1972).
[7] N. Tariq and B. O. J. Tupper, Tensor (N.S.) **31**, 42 (1977).
[8] V. Hlavatý, J. Math. Mech. **8**, 285 (1959).
[9] V. Hlavatý, J. Math. Mech. **8**, 597 (1959).
[10] V. Hlavaty, J. Math. Mech. **9**, 89 (1960).
[11] E. Ihrig, J. Math. Phys. **16**, 54 (1975).
[12] J. Plebański, Acta Phys. Pol. **26**, 963 (1964).
[13] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[14] C. B. G. McIntosh, J. Foyster, and A. W.- C. Lun, "The classification of the Ricci and Plebański tensors in general relativity using Newman–Penrose formalism", J. Math. Phys. **22**, 2620 (1981).
[15] R. Debever, R. G. McLenaghan, and N. Tariq, Gen. Relativ. Gravit. **10**, 853 (1979).
[16] M. A. H. MacCallum, "Locally isotropic space-times with non-null homogeneous hypersurfaces", preprint (1979).
[17] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **11**, 1578 (1970).
[18] W. Kundt, Z. Phys. **163**, 77 (1961).
[19] G. H. Katzin, J. Levine, and W. R. Davis, Tensor (N.S.) **21**, 51 (1970).
[20] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[21] J. Plebański, J. Math. Phys. **20**, 1946 (1979).
[22] B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[23] I. Robinson, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astron. Phys. **7**, 351 (1959).
[24] N. Tariq and B. O. J. Tupper, J. Math. Phys. **15**, 2232 (1974).
[25] A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).
[26] C. B. G. McIntosh and E. H. van Leeuwen, "Space-times admitting a vector field whose inner product with the Riemann tensor is zero," preprint, 1980.
[27] C. B. G. McIntosh and W. D. Halford, "Determination of the metric tensor from components of the Riemann tensor," preprint, 1981.

Quantization of spinor fields. II. Meaning of "bosonization" in 1 + 1 and 1 + 3 dimensions

Piotr Garbaczewski
CNRS, Marseille-Luminy, France, NORDITA, Copenhagen, Denmark, and Institute of Theoretical Physics, University of Wrocław, Wrocław, Poland [a)]

(Received 16 July 1980, accepted for publication 21 November 1980)

We demonstrate that the correspondence principle allowing us to relate the classical (c number) and quantum levels of spinor fields in 1 + 1 and 1 + 3 dimensions, involves free Bose systems with unbounded from below Hamiltonians. The necessary condition for the quantum spinor fields to be "bosonized" on the "physical" space is that for the related free Bose systems, only the non-negative part of the spectrum persists, due to constraints. Compared with the bosonization formulas, the number of independent Bose degrees of freedom necessary for a consistent formulation of the correspondence principle is doubled.

PACS numbers: 11.10.Ef

1. MOTIVATION

The "classical versus quantum" problem is far from clear for the Dirac system, both on the level of quantum field theory and relativistic quantum mechanics. In quantum mechanics the Dirac equation is believed not to admit a satisfactory classical analog, allowing at the same time an interpretation as a classical field equation for the spinor system.

It seems that here, not the relativistic form of the problem, but the spin, involves the most serious difficulties. First, spin itself is derivable from the pure Galilean background.[1] Second, it is known that the spin gives nonzero modifications to particle trajectories in the macroscopic ($\hbar \to 0$) limit, in the large-distance scale.[2] Notice also that in Q.E.D. $\hbar$ enters all quantities through $e^2/\hbar c$ so that a simple $\hbar \to 0$ rule is inapplicable and should be combined rather with the nonrelativistic limit $c \to \infty$.[3]

Basic investigations of classical analogs of the Dirac equation resulted in constructing suitable relativistic theories of spinning particles followed by their quantization.[4–14] All these attempts are based on the hope that the Dirac system can be completely understood (*and described*) *in terms of the conventional canonical variables.*

The less conventional way, though relatively simple and elegant, is to use not the usual phase space but rather a "superspace" which, in addition to the canonical variables, involves the supplementary anticommuting (Grassmann) variables, these last giving rise to spin after quantization. The appearance of pseudomechanics [15,16] provides one with a simple way to handle the relativistic quantum mechanics of the Dirac particle. Nevertheless, as pointed out in Ref. 16, the Grassmann variant of the classical mechanics cannot be applied to the real world and acquires a physical meaning *after* quantization only.

Despite its (physically) phantom nature, the generalization of the superspace concept to the continuous (field theory) level became very popular in high-energy physics, due to its calculational simplicity. If became even more popular with the advent of graded Lie algebras and supersymmetries, though these last do not need the introduction of anticommuting c numbers into the basic formalism[17,18] Likewise, in quantum mechanics, the two basic trends are met in the quantum field theory of the Dirac particle: the canonical one preferring to look for any (more or less "classical") c number level, and the Grassmann one, for which the anticommuting function ring is used to construct a pre-quantum level for the quantum field.

Though the Grassmannian way is dominant in the physical literature, there are nevertheless quite serious investigations of the c number origin of the quantum Dirac field, which date back to Klauder's paper.[19] Its idea was developed in Ref. 20, which is referred to as Paper I of the present series. Another investigation of the non-Grassmann prequantum level for spin 1/2 and Fermi lattices, together with the path integration formulas for propagators, was given in Refs. 21–23. Recall that path integrals for spinning particles were considered in Ref. 12 and quite recently in Refs. 24–27, and 28. In Ref. 25, in connection with the semiclassical quantization procedure for the continuous ferromagnetic system, the notion of a true (non-Grassmann) physical path was necessary. Then a Bose quantization of the system, under suitable constraints, was shown to conform with the well-known Bethe's solutions.

The present paper follows essentially the non-Grassmann approach, extending the earlier results of Ref. 20. Our opinion is (see, e.g., Refs. 20, 27, 29–31) that *any pseudoclassical theory described in terms of the Grassmann variables hides* (or even stronger: lacks) *its true physical content, and can in principle be reformulated as a conventional* (not pseudo) *theory of some singular canonical system.* Let us now recapitulate the basic result of Ref. 20 (i.e., I of the present series) concerning the quantization of the "classical" Dirac field. Suppose we are given free Dirac spinor fields $\overset{c}{\psi}(x), \overset{c}{\bar{\psi}}(x)$, $x \in M^4$ and let $\mathcal{F}$ be the set of all functionals $\mathbb{C} \ni \Omega(\overset{c}{\psi}, \overset{c}{\bar{\psi}})$

[a)] Permanent address.

 0022-2488/82/030442-09$02.50 © 1982 American Institute of Physics

$= \Sigma_{n,m}(\omega_{nm}, \overset{c}{\psi}^n \overset{\underline{c}}{\bar{\psi}}^m)$.

There exists in $\mathcal{F}$ a subset $\overset{c}{\mathcal{F}}$ (a prequantum level) of functionals Ω which is closed under the left multiplication operation $(*)$ of Ref. 20,

$$\overset{c}{\mathcal{F}} \ni \overset{c}{\Omega}_1 \overset{c}{\Omega}_2 \Rightarrow \overset{c}{\Omega}_1 (*) \overset{c}{\Omega}_2 = \overset{c}{\Omega}_{12} \in \overset{c}{\mathcal{F}}, \tag{1.1}$$

and for which the quantization prescription

$$\overset{c}{\Omega}(\psi,\bar{\psi}) = (\alpha|{:}\overset{c}{\Omega}(\overset{B}{\psi},\overset{\underline{B}}{\bar{\psi}}){:}|\alpha) \longrightarrow {:}\overset{c}{\Omega}(\overset{B}{\psi},\overset{\underline{B}}{\bar{\psi}}){:} \tag{1.2}$$

allows us to identify all elements of the Fermi field algebra according to

$$\mathbb{1}_F {:}\overset{c}{\Omega}(\overset{B}{\psi},\overset{\underline{B}}{\bar{\psi}}){:}\mathbb{1}_F = {:}\Omega(\psi,\bar{\psi}){:} \longleftrightarrow \overset{c}{\Omega}(\overset{c}{\psi},\overset{\underline{c}}{\bar{\psi}}). \tag{1.3}$$

Equation (1.3) is an identity in the quantum domain generated by the free quantum Dirac field $\psi(x),\bar{\psi}(x)$. The system of units is $\hbar = c = 1$ and $\overset{c}{\mathcal{F}}$ stands for a prequantum rather than classical level.

In the above $|\alpha)$ denotes a coherent state for the subsidiary CCR algebra involved in $\overset{B}{\psi},\overset{\underline{B}}{\bar{\psi}}$, these last being obtained from $\overset{c}{\psi},\overset{\underline{c}}{\bar{\psi}}$ through replacing the classical Fourier amplitudes $\overline{\alpha}_k^{\pm}(\mathbf{p}),\alpha_k^{\pm}(\mathbf{p})$, $k = 1,2$ by the Bose generators $\overset{*}{a}{}_k^{\pm}(\mathbf{p}),a_k^{\pm}(\mathbf{p})$ (the number of internal degrees of freedom is preserved while going from bosons to fermions in this construction!).

Because the fields $\psi,\bar{\psi}$ are by definition relativistic Dirac operators, the emergence of the relativistic looking objects $\overset{B}{\psi},\overset{\underline{B}}{\bar{\psi}}$ needs some explanation in light of the spin-statistics theorem: they cannot be the Wightman fields. On the other hand, the Fermi quantization (1.3), at first sight, seems to have nothing in common with any canonical quantization procedure, despite its involving bosons (for these last a canonical procedure in principle can be expected to exist).

The basic purpose of the present paper is to clarify the formal arguments of Ref. 20, by taking into account the results of Refs. 27, 29–31 and then going into the physics involved to explain the canonical quantization aspects which are inherent in (1.3), though not explicit in the formalism of Ref. 20. The concept of the Bose background for the quantum Dirac field ("bosonization", see, e.g., Ref. 29) becomes crucial at this point.The basic idea in the course of the paper is the naive version of the correspondence principle,[20] for quantum Bose systems. Take an operator expression in terms of the generators of the CCR algebra, make the so-called Bose transformation of them (translations by c-number functions), and then calculate a Fock vacuum expectation value of the result in the tree approximation [i.e., make a normal ordering before calculating $(0|\cdot|0)$]. The system of units is $\hbar = c = 1$.

Let us emphasize that the classical spinor fields $\overset{c}{\psi},\overset{\underline{c}}{\bar{\psi}}$,(1.1)–(1.3) due to the $\mathbb{1}_F(\cdot)\mathbb{1}_F$ sandwiching depend linearly on the classical amplitudes $\alpha_k^{\pm}(\mathbf{p}),\overline{\alpha}_k^{\pm}(\mathbf{p})$, $k = 1,2$. In what follows we shall admit a nonlinear dependence, which will simplify the arguments.

In Sec. 2 we demonstrate that the bosonization of the (massless) Thirring model necessitates a positivity of energy condition for the (1 + 1 dimensional) Maxwell field involved. In Sec. 3, we show that if the massive Thirring model in the charge 0 sector of the physical space is to be bosonized, then a positivity of energy condition necessarily occurs for the involved free massive neutral vector field in 1 + 1 dimensions. In Sec. 4 we prove the existence of the Maxwell-field (single potential in the Coulomb gauge) reformulation of the free quantum Dirac field, and the related algebra of observables. It appears as a consequence of the positivity of energy condition imposed on the two-potential Maxwell field Hamiltonian.

A common feature of all these cases is that a correspondence principle needs the number of Bose degrees of freedom to be doubled compared with the bosonization formulas. The underlying free field Hamiltonians have the form

$$H = H(\phi) - H(\phi'), \tag{1.4}$$

where ϕ,ϕ' are two independent free scalar (massless or massive, respectively) fields for the Thirring model, while the two independent Coulomb-gauge Maxwell fields are for the Dirac field.

In Sec. 5, we recover the two-potential Maxwell field content of the relativistic quantum mechanics of the Dirac electron. In contrast to q.f.t., a single potential formulation, seems not to be adequate here, which is inconsistent with the the Lorentz covariance properties of Dirac spinors.

2. THE MASSLESS THIRRING MODEL

A. The massless self-interacting spinor field theory in 1 + 1 dimensional space–time

$$\mathcal{L} = i\bar{\psi}\partial\!\!\!/\psi + (g/2){:}j_\mu j^\mu{:}, \quad j^\mu = \bar{\psi}\gamma^\mu\psi, \tag{2.1}$$

$$g_{00} = -g_{11} = 1, \quad \gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

is known not to possess the spinor asymptotic fields. On the other hand, if one is to follow the general principles of quantum field theory, one expects that all Heisenberg operators should be expressed in terms of some (free) asymptotic fields.

A discussion of this problem was given in Refs. 32–35, a non-Wightman neutral massless scalar field $\phi(x)$:

$$\Box\phi(x) = 0, \quad \phi^+(x)|0) = 0,$$
$$[\partial_0\phi(x^0,x^1), \quad \phi(x^0,y^1)]_- = -i\delta(x^1 - y^1) \tag{2.2}$$

was shown to represent the Thirring spinors in the form

$$\psi(x) = {:}\exp[ia\phi(x) - ib\gamma^5\tilde{\phi}(x)]{:}u. \tag{2.3}$$

$a,b \in \mathbb{R}, ab = \pi$, u being a two-component constant, the normal ordering involving an order $\{\tilde{\phi}^-,\phi^-,\phi^+,\tilde{\phi}^+\}$ of the positive and negative frequency parts of $\phi(x)$ and of the related (conjugate) field $\tilde{\phi}(x)$

$$\Box\tilde{\phi}(x) = 0, \quad \partial_\mu\tilde{\phi}(x) + \epsilon_{\mu\nu}\partial^\nu\tilde{\phi}(x) = 0,$$
$$\epsilon_{\mu\nu} = -\epsilon_{\nu\mu}, \quad \epsilon_{01} = 1. \tag{2.4}$$

The generator P_μ of space–time translations reads[33]

$$P_0 = H(\phi) = H(\tilde{\phi}) = \tfrac{1}{2}\int dx^1 :[(\partial_0\phi)^2 + (\partial_1\phi)^2]:, \tag{2.5}$$

$$P_1 = \int dx^1 :(\partial_0\phi)(\partial_1\phi):$$

and induces

$$\begin{aligned} &[\phi^{\pm}(x), P_\mu]_- = i\partial_\mu \phi^{\pm}(x), \\ &[\tilde{\phi}^{\pm}(x), P_\mu]_- = i\partial_\mu \tilde{\phi}^{\pm}(x), \\ &[\psi(x), P_\mu]_- = i\partial_\mu \psi(x). \end{aligned} \tag{2.6}$$

The canonical anticommutation relations for $\psi(x), \bar{\psi}(x)$ hold weakly on the vacuum and one-particle sectors of the (Bose field ϕ) state space, belonging to an indefinite metric carrier space.

B. Suppose we are to deal with a classical "photon" field A in 1 + 1 dimensions:

$$\Box A_\mu = 0, \quad \partial_\mu A^\mu = 0. \tag{2.7}$$

A corresponding Hamiltonian,

$$\begin{aligned} H &= \tfrac{1}{2}(\partial_\nu A_\mu)(\partial_\nu A^\mu) \\ &= \tfrac{1}{2}\{[(\partial_0 A_0)^2 + (\partial_1 A_0)^2] - [(\partial_0 A_1)^2 + (\partial_1 A_1)^2]\}, \end{aligned} \tag{2.8}$$

is obviously gauge invariant, hence allowing us to apply the Faddeev–Popov's path-integration arguments to this abelian gauge system.[36] Within the Hamiltonian formalism, the Lorentz condition $\partial_\mu A^\mu = \partial_0 A_0 - \partial_1 A_1$ appears as a constraint, which should still be accompanied by the supplementary "gauge fixing" condition, so that a canonical pair corresponding to one of the two allowed degrees $\{\pi_\mu, A_\mu\}_{\mu=0,1}$ can be completely eliminated from the formalism.

If we choose the supplementary condition in the form

$$\epsilon_{\mu\nu}F^{\mu\nu} = 0 \Rightarrow \partial_0 A_1 = \partial_1 A_0, \tag{2.9}$$

then, together with the Lorentz one, it leads to

$$\partial_\mu A_0(x) + \epsilon_{\mu\nu}\partial^\nu A_1(x) = 0, \tag{2.10}$$

which is a classical version of the definition (2.4) of the conjugate field, provided we identify $A_0 = \phi$, $A_1 = \tilde{\phi}$. Notice that (2.10) implies

$$H = H(\phi) - H(\tilde{\phi}) = 0 \Rightarrow H(\phi) = H(\tilde{\phi}). \tag{2.11}$$

With no recourse to the explicit Hamiltonian formalism, if we adopt the field (2.7) as a classical relative of the asymptotic one for the Thirring model, we can make the canonical quantization step by using a generating functional for the Thirring model Green's functions (the antisymmetry question is here left aside, see however Refs. 20 and 29)

$$\begin{aligned} W(\eta,\bar{\eta}) = \int d\mu(A) \\ \times \exp i\left\{\int d^2x[-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}] + \bar{\eta}\psi + \eta\bar{\psi}\right\}\delta(\partial_\mu A^\mu), \end{aligned} \tag{2.12}$$

where $\eta, \bar{\eta}$ are the (commuting ring!) spinor sources, and the classical (c-number) expression for ψ reads

$$\psi(A,x) = \psi(x) = \exp[ia A_0(x) - ib\gamma^5 A_1(x)]u, \quad ab = \pi. \tag{2.13}$$

Notice at this point that a formal integration with respect to A_1, provided we denote

$$A_1(x) = \int_{-\infty}^{x} d\xi\, \partial_\xi A_1(\xi), \tag{2.14}$$

replaces $A_1(x)$ in (2.12) by

$$A_1(x) \to \int_{-\infty}^{x} d\xi\, \partial_0 A_0(\xi) \tag{2.15}$$

thus giving (put $A_0 = \phi$)

$$\psi(x) = \psi(\phi,x) = \exp\left[ia\phi(x) - ib\gamma^5 \int_{-\infty}^{x} d\xi\, \dot{\phi}(\xi)\right], \tag{2.16}$$

which agrees with Mandelstam's formula, see Refs. 24, 29, and 35, if one replaces (2.16) by the normal ordered-operator expression.

3. MASSIVE THIRRING MODEL

A. Suppose we deal with a quantum massive Thirring model in the charge 0 sector[37]:

$$\begin{aligned} H = \int dx[&-i(\psi_1^*\partial_x\psi_1 - \psi_2^*\partial_x\psi_2) + m_0(\psi_1^*\psi_2 + \psi_2^*\psi_1) \\ &+ 2g_0\psi_1^*\psi_2^*\psi_2\psi_1], \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} &[\psi_i(x), \psi_j^*(y)]_+ = \delta_{ij}\delta(x-y), \\ &[\psi_i(x), \psi_j(y)]_+ = 0. \end{aligned} \tag{3.2}$$

As the analysis of the diagonalization problem for H[37] shows, the irreducibility domains for the CAR algebra (3.2) can be looked for within the general (continuous direct product) Hilbert space $\mathscr{H} = \Pi_x \otimes (h)_x$ containing the Fock state $|0\rangle$ together with a corresponding Fock irreducibility sector $\mathscr{H}(|0\rangle) \subset \mathscr{H}$

$$\psi_i(x)|0\rangle = 0, \quad \forall x \in \mathbb{R}, \quad i = 1,2. \tag{3.3}$$

The most general form of the eigenstate of H with a finite number of quasiparticles reads

$$\begin{aligned} &|\alpha_1,\ldots,\alpha_n\rangle = \int dx_1 \cdots \int dx_n \chi(x_1,\ldots,x_n) \prod_{i=1}^{n} \psi^*(x_i,\alpha_i)|0\rangle, \\ &\chi(x_1,\ldots,x_n) = \exp\left(im_0 \sum_i x_i \sinh\alpha_i\right) \\ &\qquad \times \prod_{1 \le i < j \le n}^{i=1} [1 + i\lambda(\alpha_i,\alpha_j)\epsilon(x_i - x_j)], \\ &\lambda(\alpha_i,\alpha_j) = -\tfrac{1}{2}g_0 \tanh\tfrac{1}{2}(\alpha_i - \alpha_j), \\ &\psi(x,\alpha) = \psi_1(x)\exp(\alpha/2) + \psi_2(x)\exp(-\alpha/2), \end{aligned} \tag{3.4}$$

where α can take either the value $\alpha = \beta$ or $\alpha = i\pi - \beta$ with $\tanh\beta = k/E$ (β being the rapidity of a particle with momentum k and energy E). Here

$$H(\alpha_1,\ldots,\alpha_n) = (\sum_i m_0 \cosh\alpha_i)|\alpha_1,\ldots,\alpha_n\rangle \tag{3.5}$$

and $m_0 \cosh(i\pi - \beta) = -m_0 \cosh\beta$. As a consequence, the

spectrum of H on this set of eigenstates is unbounded from below, and thus interpreted as "unphysical."

The physical part of the spectrum can however be recovered within $\mathscr{H}$, provided that we abandon the Fock sector for $\{\psi_i^*,\psi_i\}_{i=1,2}$ and use a procedure of "filling the Dirac sea" under the periodic boundary conditions for eigenstates. The energy of a state must then be measured relative to the ground state within the appropriate physical sector $\mathscr{H}_{\text{phys}} \subset \mathscr{H}$. Irrespective of the change of the sector from $\mathscr{H}(|0\rangle)$ to $\mathscr{H}_{\text{phys}}$, all observables of the system can be viewed as bounded (in $\mathscr{H}(|0\rangle)$ or $\mathscr{H}_{\text{phys}}$, respectively) functions of the fundamental fields $\psi_i^*(x)$, $\psi_i(x)$, $i = 1,2$, where creation and annihilation operators occur due to

$$b_i(k) = \int \frac{dx}{(2\pi)^{1/2}} \exp(-ikx)\,\psi_i(x). \tag{3.6}$$

Notice that in the absence of the self-coupling ($g_0 = 0$), after a canonical transformation

$$\begin{aligned} B_1(k) &= \cos\theta(k)b_1(k) + \sin\theta(k)b_2(k), \\ B_2(k) &= -\sin\theta(k)b_1(k) + \cos\theta(k)b_2(k), \\ \tan 2\theta(k) &= m_0/k, \end{aligned} \tag{3.7}$$

(3.1) converts into

$$H_0 = \int dk\,(k^2 + m_0^2)^{1/2}(B_1^*(k)B_1(k) - B_2^*(k)B_2(k)), \tag{3.8}$$

where under $B_i(k)|0\rangle = 0\ \forall i,k$ the spectrum of H_0 is obviously nonpositive in $\mathscr{H}(|0\rangle)$.

B. Let us introduce a neutral massive vector field $U_\mu(x)$ in 1 + 1 dimensions (which is not a Proca one, unless one imposes the subsidiary condition):

$$(\Box + m_0^2)U_\mu(x) = 0. \tag{3.9}$$

Its Hamiltonian is

$$\begin{aligned} H &= \tfrac{1}{2}\{\partial_\nu U_\mu \partial_\nu U^\mu + m_0^2 U_\mu U^\mu\} \\ &= \tfrac{1}{2}[(\partial_\nu U_0)^2 + m_0^2 U_0] - \tfrac{1}{2}[(\partial_\nu U_1)^2 + m_0^2 U_1] \\ &= \int dk\,(k^2 + m_0^2)^{1/2}[a_0^*(k)a_0(k) - a_1^*(k)a_1(k)], \end{aligned} \tag{3.10}$$

where

$$\begin{aligned} &[a_i(k),a_j^*(p)]_- = \delta_{ij}\delta(k-p), \\ &[a_i(k),a_j(p)]_- = 0, \\ &a_i(k)|0\rangle = 0, \quad \forall i,k, \quad i = 0,1, \end{aligned} \tag{3.11}$$

thus constituting a Fock representation of the CCR algebra. By virtue of Ref. 38, we can identify the Fock vacuum $|0\rangle$ with that of the free Thirring model, as in Refs. 20 and 38, the Fermi generators $B_i^*(k)$, $B_i(k)$ can be completely given in terms of the Bose generators $a_i^*(k)$, $a_i(k)$ (3.11). The coincidence of the number (two) of internal degrees of freedom is crucial in this construction of the CAR within the CCR algebra. Then $1_F H 1_F = H_0$, where 1_F is the Fermi operator unit, H is (3.10), while H_0 is (3.8), see Refs. 20 and 29.

This means that all functions $\Omega(\psi^*,\psi)$ can be rewritten as functions $\Omega(\psi^*,\psi) = F(B^*,B) = G(a^*,a)$. Hence all Fermi field observables can be given as observables of the system of two independent neutral scalar fields with the same mass m.

Let us emphasize that the spectrum of H, (3.10) analogously to that of H_0 (3.8), is unbounded from below within the Fock space. However, for the Proca field a subsidiary condition

$$\partial_\mu U^\mu = 0 \tag{3.12}$$

would remove this unboundedness difficulty, as in Refs. 32 and 33; then

$$\begin{aligned} &U_\mu = -m_0^2 \epsilon_{\mu\nu}\partial^\nu U, \\ &U := \epsilon^{\mu\nu}\partial_\mu U_\nu, \\ &(\Box + m_0^2)U = 0, \end{aligned} \tag{3.13}$$

which proves that a system $\{(\Box + m_0^2)U_\mu = 0 = \partial_\mu U_\mu\}$ is equivalent to a neutral massive scalar field. The restriction $\partial_\mu U^\mu = 0$ makes the spectrum of H positive definite, at the same time reducing a two-component system to a single-component one.

C. Let us here comment that the very same positivity requirement for H, (3.1) makes the massive Thirring model equivalent to the quantum sine-Gordon system (within suitable limitations on the coupling constants values).

As found in Refs. 31 and 39, both classical and quantum sine-Gordon fields (including solitons) do exhibit a neutral massive free-field structure, hence the field U can be quite naturally embedded in the sine-Gordon framework. We conjecture that a positivity condition for U_μ induces a positivity condition for (3.1).

D. On the other hand, by using a boson transformation concept[39] (generalized coherent states come into account here), Thirring model observables

$$\Omega(\psi^*,\psi) = G(a^*,a), \tag{3.14}$$

if Bose transformed,

$$G(a^*,a) \to G(a^* + \bar{\alpha}, a + \alpha), \tag{3.15}$$

give rise in the tree approximation to the following (Fock) vacuum expectation values:

$$\begin{aligned} &\langle 0|G(a^* + \bar{\alpha}, a + \alpha)|0\rangle \to \langle 0|:G(a^* + \bar{\alpha}, a + \alpha):|0\rangle = G(\bar{\alpha},\alpha), \\ &\langle 0|:\psi(a^* + \bar{\alpha}, a + \alpha):|0\rangle = \overset{c}{\psi}(\bar{\alpha},\alpha). \end{aligned} \tag{3.16}$$

By exploiting this procedure, the quantum Thirring model Hamiltonian goes over to the classical Thirring model Hamiltonian of exactly the same form, with $\overset{c}{\psi} = \overset{c}{\psi}(\bar{\alpha},\alpha)$. In the above, :.: denotes a normal ordering with respect to Bose variables.

The classical Thirring model is known[40] to be a completely integrable system, whose classical spectrum reads

$$H = \int_0^\infty [\rho_1(s) - \rho_2(s)][p^2(s) + m_0^2]^{1/2} ds^2 + \sum_{n=1}^{N} (A_n^2 + M_n^2)^{1/2}, \tag{3.17}$$

where momenta of mass m_0 particles with densities ρ_1, ρ_2 are given by $p(s) = (m_0/2)(s^{-2} + s^2)$. The remaining (discrete) part of the spectrum is due to solitons.

Hence the two neutral scalar fields are necessarily present on the classical Thirring level: the procedure (3.14)–(3.16) does require the two scalar fields to define $G(\bar{\alpha},\alpha)$ while for the sine-Gordon system, the same procedure[39] would recover a single neutral field structure (which is consistent with the classical spectral solution of Ref. 41).

E. Our conclusion is that the two, entirely different classical, completely integrable models, i.e., the massive Thirring and sine-Gordon systems, both exhibit the neutral free field structure, whose respective quantum images can coincide (Coleman's equivalence) provided a positivity condition for U_μ induces a positivity condition on (3.1).

This removes the redundant degree of freedom from the theory, thus replacing a two-component Bose-field formulation, which is characteristic for the Thirring model, by a single scalar-field formulation, corresponding to the sine-Gordon system. *Warning*: In contrast to the massless Thirring model, the underlying massive free scalars are not asymptotic fields at all. Nevertheless, all quantum and classical observables of the Thirring and sine-Gordon systems can be completely expressed in terms of them. For a review of analogous (quasiparticle) structures, see, e.g., Ref. 29.

4. ELECTROMAGNETIC (FREE FIELD) STRUCTURE OF THE QUANTIZED DIRAC FIELD

A. Previous 1 + 1 dimensional considerations can be summarized in the shorthand notions of the "free massless neutral field structure of the Thirring model" and the "free massive neutral field structure of the massive (and the related sine-Gordon) model."

Needless to say, in both massive and massless Schwinger models (quantum electrodynamics in 1 + 1 dimensions) no free fermions might occur in the asymptotic particle spectrum.

For the massless model, if it is provided with a subsidiary condition (to guarantee positivity of the spectrum on the physical subspace of the general indefinite metric Hilbert space) the only field of importance remains the free Proca field $U_\mu(x)$ (and hence a massive scalar U).[32,42] In the case of the massive model the spectrum consists of the neutral massive bosons identified with those of the massive sine-Gordon system (no definite free field structure of it is known to me, but it would surely be a scalar Bose one).

B. Our wisdom about quantum electrodynamics in 1 + 3 dimensions and hence the quantized Dirac and electromagnetic fields, follows from Refs. 43 and 44.

We shall work with the so-called scattering representations of the electromagnetic (free, asymptotic, in–out) field algebra; they have an energy momentum operator satisfying the relativistic spectrum conditions and can be defined in the charged sectors of the physical Q.E.D. Hilbert space. With an appropriate definition of the charge operator Q in $\mathscr{H}_{\text{phys}}$ following from $\Box A_\mu(x) = j_\mu(x)$ with $j_\mu(x)$ interpreted as an electric current, one can prove that a Hilbert space $\mathscr{H}_{\text{in}} \subset \mathscr{H}_{\text{phys}}$ is in the domain of Q and $Q\mathscr{H}_{\text{in}} = 0$. Then one[44] proves that no asymptotic (free) charged field ψ_{in} can exist in $\mathscr{H}_{\text{phys}}$, which is local with respect to $F^{\text{in}}_{\mu\nu}$. A scattering representation algebra we denote $\pi(\mathscr{A})$.

An energy operator P_μ of $\pi(\mathscr{A})$ can be decomposed into $P_\mu = P_{\mu\,\text{as}} + P_{\mu\,\text{ch}}$ where $P_{\mu\,\text{as}}$ is associated with $\pi(\mathscr{A})''$ and hence describes the energy momentum of the asymptotic electromagnetic field configuration. $P_{\mu\,\text{ch}}$ is associated with $\pi(\mathscr{A})'$ and thus describes charges and fields without electromagnetic interactions.[44] Here an important relation holds true:

$$Sp\, P_{\mu\,\text{ch}} \subseteq Sp P_{\mu\,\text{as}} = \bar{V}^+, \tag{4.1}$$

so that the spectrum of charges, in principle, can be completely recovered within the spectrum of the asymptotic electromagnetic field configuration. Compare here, e.g., also Refs. 29–31, where quite analogous conclusions were drawn in our studies of the Bose–Fermi "metamorphosis."

Moreover, the charged (infra) states do necessarily generate non-Fock irreducibility sectors of the asymptotic field algebra; this conforms well with the traditional infinite direct product constructions, where the Hilbert space $\mathscr{H} = \Pi_x \otimes (h)_x$ carries a reducible representation of the field algebra $\mathscr{A}$, and one must specify the generating vector to select a definite irreducibility sector for $\mathscr{A}$ and thus to specify $\pi(\mathscr{A})$. One should also know that only the radiation field associated with the charge 0 sector of $\mathscr{H}_{\text{phys}}$ is Lorentz covariant.

For the physics, the following is essential.[44] The probability distributions of the momenta of the asymptotic charged infraparticles in a scattering state can, in principle, be determined by measurements of the asymptotic electromagnetic field alone. This suggests that even though the charges in Q.E.D. are not confined (in contrast to 1 + 1 dimensional theories) the whole Dirac theory should admit a reconstruction in terms of the asymptotic free electromagnetic fields.

C. In Ref. 20, we constructed elements of the Dirac field algebra in terms of free Fermi fields. The correspondence principle leading to a non-Grassmann (commuting ring of spinors) Dirac level, involved there a CCR algebra, with the same number of internal degrees of freedom, as here of the CAR, namely,

$$\{a_i^*,a_i,|0)\}_{i=1,2,3,4}\to\{b_i^*,b_i,|0)\}_{i=1,2,3,4},$$
$$[a_i(\mathbf{p}),a_j^*(\mathbf{k})]_- = \delta_{ij}\delta(\mathbf{k}-\mathbf{p}),$$
$$[b_i(\mathbf{p}),b_j^*(\mathbf{k})]_+ = \delta_{ij}\delta(\mathbf{k}-\mathbf{p}),$$
$$[a_i(\mathbf{p}),a_j(\mathbf{k})]_- = 0, \quad [b_i(\mathbf{p}),b_j(\mathbf{k})]_+ = 0,$$
$$a_i^*(\mathbf{k})|0) = b_i^*(\mathbf{k})|0),$$
$$a_i(\mathbf{k})|0) = b_i(\mathbf{k})|0) = 0,\forall,i,\mathbf{k}, \tag{4.2}$$

see Ref. 38. It was the main cause of problems with a physical interpretation of the underlying bosonization in 1 + 3 dimensions.

D. However, the lesson of 1 + 1 dimensional models suggests that this seemingly unphysical Bose background of the quantum Dirac system should be converted into a physical one if suitable domain constraints would enter. We claim that this role is played again by the positivity-of-energy condition.

A classical Maxwell field can be described in terms of the two independent four potentials,[45,46] M_μ, N_μ so that

$$\partial_\nu F_{\mu\nu} = j_\mu, \quad \partial_\nu \tilde{F}_{\mu\nu} = 0,$$
$$F_{\mu\nu} = M_{\mu\nu} - \tilde{N}_{\mu\nu} = \partial_\mu M_\nu - \partial_\nu M_\mu - \epsilon_{\mu\nu\alpha\beta}\partial^\alpha N^\beta, \tag{4.3}$$
$$\tilde{F}_{\mu\nu} = N_{\mu\nu} + \tilde{M}_{\mu\nu} = \partial_\mu N_\nu - \partial_\nu N_\mu + \epsilon_{\mu\nu\alpha\beta}\partial^\alpha M^\beta,$$

and the gauge freedom is significantly enlarged

$$M_\mu \to M_\mu + \partial_\mu \Lambda(x), \quad N_\mu \to N_\mu + \partial_\mu \chi(x),$$
$$M_\mu \to M_\mu^0 + M_\mu, \quad N_\mu \to N_\mu^0 + N_\mu, \tag{4.4}$$
$$\partial_\mu M_\nu^0 - \partial_\nu M_\mu^0 - \epsilon_{\mu\nu\alpha\beta}\partial^\alpha N^{0\beta} = 0.$$

Consequently, by a proper choice of gauge ($N_\mu^0 = -N_\mu$ plus the Lorentz or Coulomb one), we can remove the redundant degrees of freedom, thus reducing the problem to a single potential one. This is the case for sourceless and electric examples. In the presence of magnetic sources the second potential cannot be eliminated.

Obviously, due to the gauge freedom, the Lorentz condition can be imposed on both potentials: $\partial_\mu M^\mu = 0 = \partial_\mu N^\mu$ and the second kind of gauge freedom still allows us to make an appropriate (like, e.g., the Coulomb one) gauge choice for the system.

The Lagrangian for the two-potential system (4.3) can be introduced according to convention

$$\begin{aligned}\mathcal{L} &= -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}\\ &= -\tfrac{1}{4}M_{\mu\nu}M^{\mu\nu} + \tfrac{1}{4}N_{\mu\nu}N^{\mu\nu} + \partial^\mu(\epsilon_{\mu\nu\rho\sigma}\partial^\rho M^\sigma N^\nu)\\ &= \mathcal{L}_M - \mathcal{L}_N + \partial^\mu(\epsilon_{\mu\nu\rho\sigma}\partial^\rho M^\sigma N^\nu).\end{aligned} \tag{4.5}$$

The constraints

$$\partial_0 N_0 = \partial_0 M_0 = 0 = \partial_\mu M^\mu = \partial_\nu N^\nu \tag{4.6}$$

reduce both potentials to the radiation gauge. In the presence of one more constraint

$$\partial_\mu(\epsilon_{\mu\nu\rho\sigma}\partial^\rho M^\sigma N^\nu) = 0, \tag{4.7}$$

the Lagrangian $\mathcal{L}$ can be replaced by $\mathcal{L} = \mathcal{L}_M - \mathcal{L}_N$, which is quite analogous to the Lagrangians (2.8) and (3.7) of the previous sections. Then the constraint

$$M_\mu N^\mu = 0 \tag{4.8}$$

allows a complete elimination of one of the two potentials, leaving us in (4.5) with a single potential in the radiation gauge; notice that both (4.7) and (4.8) are consistent with the demand $N_\mu = 0\ \forall\mu$, and because the gauge freedom is removed from the system, we conclude that $\mathcal{L} = \mathcal{L}_M - \mathcal{L}_N$ if supplied with six constraints (4.6)–(4.8) becomes a single potential system in the radiation gauge. In the path integral framework this observation can be compactly written as follows[36]:

$$\begin{aligned}&\int \exp\{iS[A]\}\prod_x \delta(\partial_\mu A^\mu)\delta(\partial_0 A^0)\prod_\mu dA_\mu(x)\\ &= \int\int \exp\{iS[M,N]\}\prod_x \delta(\partial_\mu M^\mu)\\ &\quad\times\delta(\partial_0 M^0)\delta(\partial_\nu N^\nu)\delta(\partial_0 N^0)\\ &\quad\times\delta(M_\mu N^\mu)\delta(\partial^\mu\epsilon_{\mu\nu\rho\sigma}\partial^\rho M^\sigma N^\nu)\\ &\quad\times\prod_\mu dM_\mu(x)\prod_\nu dN_\nu(x),\end{aligned} \tag{4.9}$$

with

$$S[M,N] = \int d^4x\{\mathcal{L}(M) - \mathcal{L}(N)\},$$
$$S[A] = \int d^4x\,\mathcal{L}(A).$$

E. Let us emphasize that the constraints (4.7) and (4.8) play the same role as the constraints (2.10) and (3.12) in cases of the massless and massive Thirring models, respectively; they transform an involved quantum Bose system with a nonpositive spectrum into a manifestly positive one, but at the price of diminishing the number of (Bose) degrees of freedom. Without the "mixing" conditions (4.7) and (4.8) we deal with a doubled Maxwell field in the radiation gauge, whose quantized Hamiltonian can be equivalently written as

$$\begin{aligned}H &= H(M) - H(N)\\ &= \int d^3k\,|\mathbf{k}|\sum_{\lambda=1}^{2}[a_M^*(\mathbf{k},\lambda)a_M(\mathbf{k},\lambda)\\ &\quad - a_N^*(\mathbf{k},\lambda)a_N(\mathbf{k},\lambda)],\end{aligned} \tag{4.10}$$

where

$$\begin{aligned}H(N) &= \tfrac{1}{2}\int d^3x{:}\mathbf{N}^2 + (\nabla\times\mathbf{N})^2{:}\\ &= \int d^3k\,|\mathbf{k}|\sum_{\lambda=1}^{2}a_N^*(\mathbf{k},\lambda)a_N(\mathbf{k},\lambda),\end{aligned} \tag{4.11}$$

$$[a_N(\mathbf{k},\lambda),a_N^*(\mathbf{k}',\lambda')]_- = \delta_{\lambda\lambda'}\cdot\delta(\mathbf{k}-\mathbf{k}'),$$
$$[a_N(\mathbf{k},\lambda),a_N(\mathbf{k}',\lambda')]_- = 0,$$
$$[a_N^\#(\mathbf{k},\lambda),a_M^\#(\mathbf{k}',\lambda')]_- = 0,$$
$$a_N(\mathbf{k},\lambda)|0) = a_M(\mathbf{k},\lambda)|0) = 0,$$
$$a_N(\mathbf{k},\lambda) = i\int d^3x\,\exp(i\mathbf{k}\mathbf{x})\overleftrightarrow{\partial}_0\epsilon(\mathbf{k},\lambda)\times\mathbf{A}(x)/[2|\mathbf{k}|(2\pi)^3]^{1/2},$$

with $\{\epsilon(\mathbf{k},\lambda),\mathbf{k}/|\mathbf{k}|\}_{\lambda=1,2}$ forming a basis system in E^3. Quantally, the constraints (4.7) and (4.8) become the domin-

ant ones within an infinite (continuous) direct product space carrying a reducible representation of the CCR algebra (4.11). The metric in $\mathscr{H}$ is indefinite.[44]

We identify the quartet of Bose generators

$$\begin{aligned} &a_M(\mathbf{k},1)=a_1(\mathbf{k}), \quad a_M(\mathbf{k},2)=a_2(\mathbf{k}),\\ &a_N(\mathbf{k},1)=a_3(\mathbf{k}), \quad a_N(\mathbf{k},2)=a_4(\mathbf{k}),\\ &[a_k(\mathbf{p}),a_l^*(\mathbf{q})]_-=\delta_{ki}\delta(\mathbf{p}-\mathbf{q}),\\ &[a_k(\mathbf{p}),a_i(\mathbf{q})]_-=0,\\ &a_k(\mathbf{p})|0)=0 \quad \forall\mathbf{p}, \quad k=1,2,3,4, \end{aligned} \tag{4.12}$$

with the one, (4.2), used in the Bose construction of Fermi generators for the quantum Dirac field[38,20] in 1 + 3 dimensions.

The previous analysis shows that the Dirac operators $\psi=\psi(a^*,a)$, $\bar{\psi}=\bar{\psi}(a^*,a)$ which are thus completely given in the two-potential Maxwell framework, after imposing the quantum images of constraints (4.7) and (4.8) become equivalent to the corresponding single-potential operators $\psi(A)$, $\bar{\psi}(A)$ with A_μ in the radiation gauge. It conforms well with Luther's observation,[47] that a two-component Bose system (field) should suffice to generate a free Dirac field in 1 + 3 dimensions (both in the massive and massless cases).

F. To construct a classical (c-number) analog of the quantum Dirac field, see, e.g., Ref. 20, one has to work in the two-potential Maxwell framework, so that the conventional,[48] Hamiltonian density,

$$H=\psi^*(+\alpha\nabla+\beta m)\psi, \quad \psi=\psi(a^*,a), \tag{4.13}$$

after making the boson transformation[39]: $a^*\to a^*+\bar{\gamma}$, $a\to a+\gamma$ and then taking the Fock vacuum expectation value in the tree approximation (with respect to the Bose degrees), leads to

$$(0|:H(a^*+\bar{\gamma},a+\gamma):_B|0)=H(\bar{\gamma},\gamma)=\overset{c}{\psi}{}^*(\alpha\nabla+\beta m)\overset{c}{\psi}, \tag{4.14}$$

where $\overset{c}{\psi}$ is a classical (c-number) Dirac field, $:\,:_B$ designating a normal ordering of Bose creation and annihilation operators $\{a_i^*(\mathbf{k}),a_i(\mathbf{k})\}_{i=1,2,3,4}$. Here

$$\overset{c}{\psi}(\bar{\gamma},\gamma)=(0|:\psi(a^*+\bar{\gamma},a+\gamma):_B|0), \tag{4.15}$$

which significantly differs from the definition of $\overset{c}{\psi}$ given in Ref. 20, by being nonlinear in $\bar{\gamma},\gamma$.

Remark 1: Apart of the manifest Bose background, both classically and quantally, Dirac fields preserve their canonical identity as separate (from the involved Maxwell fields) objects. Quite an analogous property was observed in Ref. 39 for the sine-Gordon system: quantum and classical solitons exhibit a free (massive neutral) field structure, but within the canonical formalism they can be viewed independently of the underlying free fields.

Remark 2: The Hamiltonian (4.13) is not positive definite and the normal ordering $:\,:_F$ with respect to the four Fermi generators $\{b_i(a^*,a,\mathbf{k}),b_i(a^*,a,\mathbf{k})\}_{i=1,\ldots,4}$ is known to convert (4.13) into a conventional positive definite operator $:H:_F=H_{\text{Dirac}}$. This Hamiltonian should be identified with $P_{\mu\,\text{ch}}$ in Ref. 44

In the path-integral framework, the generating functional for spinor Green's functions of the Dirac field reads then (compare, e.g., also Ref. 20)

$$\begin{aligned} W(\bar{\eta},\eta)=\int\int \exp\left\{i\int d^4x\left[\overset{c}{\bar{\psi}}(i\partial+m)\overset{c}{\psi}+\bar{\eta}\psi+\eta\bar{\psi}\right]\right\}\\ \times\prod_x \delta(\partial_\mu M^\mu)\delta(\partial_0 M^0)\delta(\partial_\nu N^\nu)\delta(\partial_0 N^0)\\ \times\delta(M_\mu N^\mu)\delta(\partial^\mu(\epsilon_{\mu\nu\rho\sigma}\partial^\rho M^\sigma N^\nu))\\ \times\prod_\mu dM_\mu(x)\prod_\nu dN_\nu(x), \end{aligned} \tag{4.16}$$

where

$$\overset{c}{\mathscr{L}}=\overset{c}{\bar{\psi}}(i\partial+m)\overset{c}{\psi}=\overset{c}{\pi}\overset{c}{\dot{\psi}}-\overset{c}{H}, \quad \overset{c}{\pi}=i\overset{c}{\psi}{}^* \tag{4.17}$$

and $\overset{c}{\psi},\overset{c}{H}$ are given by (4.14) and (4.15).

By taking into account the constraints it is possible to perform the four integrations with respect to $\Pi_\mu\, dN_\mu(x)$ so that we are left with a single potential radiation gauge for $\overset{c}{\mathscr{L}},\overset{c}{\psi}$.

G. Notice that because of the Coulomb gauge adopted, there is no manifest Lorentz invariance in either (4.16) or (4.9). However, Lorentz covariance properties are correct, as can be most easily seen in the single potential-radiation gauge framework for Dirac operators.

Let $\Lambda\in\mathscr{L}^\uparrow_+$, then the Maxwell field transforms as follows:

$$F^\Lambda_{\mu\nu}(x)=\tau_\Lambda(F_{\mu\nu}(\Lambda x))=\Lambda^{\ \alpha}_{\mu}\Lambda^{\ \beta}_{\nu}F_{\alpha\beta}(x), \tag{4.18}$$

while the Dirac field transforms according to

$$\begin{aligned} \psi_\Lambda(x)=\sigma_\Lambda(\psi(\Lambda x))=S^{-1}(\Lambda)\psi(\Lambda x)\\ \times=U(\Lambda)\psi(x)U(\Lambda)^{-1}, \end{aligned} \tag{4.19}$$

with $S(\Lambda)$ a finite-dimensional representative of Λ acting on spinor indices. Here,

$$S^{-1}\gamma^\mu S=\Lambda^{\ \mu}_{\nu}\gamma^\nu \tag{4.20}$$

and γ^μ is the Dirac matrix.

If, according to Sec. 4F, we adopt a single potential Coulomb gauge structure $\psi=\psi(A)$ of the quantized field ψ, then a Lorentz transformation $U(\Lambda)$ must be accompanied by the Lorentz transformation $V(\Lambda)$ of the Maxwell potential, induced by (4.18):

$$A'^\mu_\Lambda(x)=V(\Lambda)A^\mu(x)V(\Lambda)^{-1}=\Lambda^{\ \mu}_{\nu}A^\nu(\Lambda x). \tag{4.21}$$

However, the Coulomb gauge does not persist in the transformation unless an accompanying gauge transformation $W(\alpha,\Lambda)$ is performed on A'^μ_Λ to restore the gauge:

$$A^\mu_\Lambda(x)=W(\alpha,\Lambda)A'^\mu_\Lambda(x)W(\alpha,\Lambda)^{-1}=A'^\mu_\Lambda-\partial^\mu\alpha_\Lambda(\Lambda x). \tag{4.22}$$

α_Λ is an operator-valued gauge "parameter."

This means that the Lorentz mapping $U(\Lambda)$ of free Dirac operators induces (and inversely is induced by) the following mapping in the asymptotic electromagnetic field algebra:

$$\Lambda:\psi\to\psi_\Lambda \Longleftrightarrow \psi(\Lambda)\to\psi(\Lambda_\Lambda),$$
$$A^\mu_\Lambda(x) = W(\alpha,\Lambda)V(\Lambda)A^\mu(x)V(\Lambda)^{-1}W(\alpha,\Lambda)^{-1}. \tag{4.23}$$

Let us mention that (charged) generating vectors for irreducible representations of the scattering algebra $\pi(\mathscr{A})$ transform covariantly under the Lorentz group in consistency with (4.23).[44]

5. ASYMPTOTIC MAXWELL FIELD CONTENT OF THE FREE DIRAC SYSTEM, CONTINUED: RELATIVISTIC QUANTUM MECHANICS OF THE DIRAC ELECTRON

A. Assume that we work within a two-potential framework (4.10) so that, due to the Coulomb gauge, a quartet of canonical Bose operators $\{a_i^*(\mathbf{k}),a_i(\mathbf{k})\}_{i=1,2,3,4}$ (4.12) emerges, with the N_μ, M_μ dependence fixed by (4.11).

Let $f(\mathbf{p})$ be a normalized momentum space function

$$\int \frac{d^3k}{|\mathbf{k}|}\bar{f}(\mathbf{k})f(\mathbf{k}) = 1. \tag{5.1}$$

We introduce

$$\int \frac{d^3k}{|\mathbf{k}|^{1/2}} a_i(\mathbf{k})\bar{f}(\mathbf{k}) =: a_i^f, \quad i = 1,2,3,4$$
$$[a_i^f,a_k^{f*}]_- = \delta_{ik}, \quad [a_i^f,a_k^f]_- = 0, \tag{5.2}$$

where the manifest f dependence $a_i = a_i^f$ occurs. We shall fix the choice of $f = f(\mathbf{p})$ and the f index will be omitted for simplicity.

With the CCR algebra generators (5.2), provided we take a Fock representation: $a_i|0) = 0\ \forall i = 1,2,3,4$, we can construct the new operators $\{\mathbf{s},\boldsymbol{\xi}\}$:

$$s_k - \xi_k = i\epsilon_{ijk}a_j^*a_k, \quad s_k + \xi_k = i(a_k^*a_4 - a_4^*a_k). \tag{5.3}$$

They satisfy the following commutation relations[27]:

$$[s_i,N]_- = 0 = [\xi_i,N]_- = [N,\mathbf{s}^2]_-, \tag{5.4}$$
$$N = \sum_{i=1}^4 a_i^*a_i, \quad \mathbf{s}^2 = \boldsymbol{\xi}^2,$$

and in addition, the SU(2)×SU(2) ones:

$$[s_i,\xi_i]_- = 0,$$
$$[s_i,s_j]_- = i\epsilon_{ijk}s_k, \quad [\xi_i,\xi_k]_- = i\epsilon_{ijk}\xi_k. \tag{5.5}$$

All operator identities are valid in the Hilbert space $h = h_f$ constructed from the Maxwell field Fock vacuum $|0)$ by applying $\{a_i^*,a_i\}_{i=1,2,3,4}$.

As shown in Ref. 27, by using $\mathbf{s},\boldsymbol{\xi}$ and the quantum mechanical momentum operator $\mathbf{p} = \{p_k = -i\partial/\partial x_k\}_{k=1,2,3}$, the Hamiltonian

$$H = 2m\xi_3 + 4\xi_1(\mathbf{s}\cdot\mathbf{p}), \tag{5.6}$$

if provided with a domain restriction

$$(N-1)|\phi\rangle = 0 \tag{5.7}$$

in the Hilbert space $h\otimes\mathscr{H}$ ($\mathscr{H}$ carrying a Schrödinger representation of the CCR algebra $\{x_i,p_i\}_{i=1,2,3}$) becomes equivalent to Dahl's Hamiltonian[14]:

$$H_F = 2m\xi_{F3} + 4\xi_{F1}(\mathbf{s}_F\cdot\mathbf{p}) \tag{5.8}$$

Here, operators

$$H_F = P_{1/2}HP_{1/2}, \quad \mathbf{s}_F = P_{1/2}\mathbf{s}P_{1/2}, \quad \boldsymbol{\xi}_F = P_{1/2}\boldsymbol{\xi}P_{1/2}, \tag{5.9}$$
$$P_{1/2} = \prod_{k=1}^4 \{:\exp(-a_k^*a_k):+a_k^*:\exp(-a_k^*a_k):a_k\} - \prod_{k=1}^4 :\exp(-a_k^*a_k):$$

act invariantly in $h_{1/2}\otimes\mathscr{H}$, $h_{1/2} = P_{1/2}h$.

The matrix realization of H_F in $h_{1/2}\otimes\mathscr{H}$ coincides with the well-known Dirac expression

$$H_F\psi = (m\beta + \boldsymbol{\alpha}\cdot\mathbf{p})\psi = i\,\partial\psi/\partial t, \tag{5.10}$$

provided ψ is a bispinor composed of the expansion coefficients of the generalized vector $|\phi\rangle = |\phi,\mathbf{x},t\rangle = \Sigma_{k=1}^4\psi_k(\mathbf{x},t)|k\rangle$ in the basis of $h_{1/2}$: $|k\rangle = a_k^*|0)$.

B. If we smear $|\phi,\mathbf{x},t\rangle$ with a space-time dependent test function $g = g(\mathbf{x},t)$

$$|g\cdot\phi\rangle = \int d^4x\, g(x)|\phi,x\rangle = \sum_{i=1}^4 \psi_i(g)|i\rangle, \tag{5.11}$$

then the only effect of the Lorentz transformation on $|g\cdot\phi\rangle$ following from the Lorentz invariance of (5.10), is due to the unitary mapping $U(\Lambda)$ inducing a base change in $h_{1/2}$

$$U(\Lambda)|g\cdot\phi\rangle = |g\cdot\phi\rangle_\Lambda = \sum_{k=1}^4 \psi_k(g)|k,\Lambda\rangle, \tag{5.12}$$

with

$$|k,\Lambda\rangle = \sum_{i=1}^4 T(\Lambda)_{kl}|l\rangle \tag{5.13}$$

such that [compare, e.g., (4.20)]

$$|g\cdot\phi\rangle_\Lambda = \sum_{k=1}^4 \psi_k^\Lambda(g)|k\rangle = \sum_{k=1}^4\sum_{l=1}^4 (S(\Lambda)_{kl}^{-1}\psi_l)|k\rangle. \tag{5.14}$$

$S(\Lambda)$ is a finite dimensional representative of Λ in E^4. Equation (5.13) allows a nontrivial mixing of the Maxwell degrees of freedom, which preserves the Coulomb gauge of both Maxwell potentials after a Lorentz mapping. Notice, however, that the correct spinor transformation law under Λ can here be generated by supplementing the conventional Lorentz transformations of M_μ, N_μ with the appropriate "mixing" gauge transformations plus the ones necessary to restore the Coulomb gauge [like (4.22)]. Notice that the constraint (5.7) still does not remove the gauge freedom.

Remark 1: In a fixed Lorentz frame, the above quantization procedure for the Dirac system can be realized in the formal path integration framework[27]: $\hbar = c = 1$

$$\bar{\gamma}_k = (1/\sqrt{2})(\rho_k - i\pi_k),$$
$$\gamma_k = (1/\sqrt{2})(\rho_k + i\pi_k), \quad k = 1,2,3,4$$
$$Z = \int \prod_{k=1}^4 \mathscr{D}\rho_k\mathscr{D}\pi_k \prod_{i=1}^3 \mathscr{D}p_i\mathscr{D}x_i \prod_t \delta(\rho^2+\pi^2-2) \times\delta((\rho,\pi))(\pi^2-\rho^2)\exp\left[i\int dt\left\{\sum_{k=1}^4 \pi_k\dot{\rho}_k + \sum_{i=1}^3 p_i\dot{x}_i - H_{c1}(\rho,\pi,\mathbf{p})\right\}\right],$$
$$(\rho,\pi) = \sum_{k=1}^4 \rho_k\pi_k, \tag{5.15}$$

where the "classical" Hamiltonian follows from the quantized one: $H_F = P_{1/2} H P_{1/2}, [P_{1/2}, H]_- = 0$ by making the boson transformation of all Bose creation–annihilation operators and then calculating the Fock vacuum expectation value of the obtained quantities in the tree approximation:

$$H_{c1}(\rho,\pi,\mathbf{p}) = (0|{:}H(a^* + \bar{\gamma}, a + \gamma, A^* + \bar{\beta}, A + \beta){:}_B|0).$$

Originally, we used in Ref. 27 an incorrect form of the action which in addition to H_{c1} included more terms linear in momentum.)

Remark 2: Recall that if we consider the quantum field theory version of the Dirac system, the single particle Hamiltonian (4.10) is put in between the quantized Dirac fields, in its matrix form. Hence the whole Maxwell content of a single particle theory is lost.

Remark 3: In contrast to quantum field theory, it seems that the internal space of the Dirac particle needs a two-potential Maxwell framework rather than a single-potential one, at least to preseve the Lorentz covariance properties of the whole procedure. Notice that the impossibility of removing one potential among the two is usually connected with the presence of magnetic charges in the system.

Remark 4: The J. Math. Phys. referee has acquainted me with some papers which are relevant to this paper. I would like to point out the DeFacio, Hammer, and Tucker approach[49,50] to the quantization of relativistic fields: The equation of motion for the field and the resulting conserved current are the only data needed to derive all the necessary commutators or anticommutators. No *a priori* need for the canonical formalism appears in this construction of arbitrary spin electrodynamics. Application to the Dirac system which is minimally coupled to the vector meson field is of major importance for us.

The existence of a classical c-number set of solutions to the field equations is absolutely required for the existence of a quantized theory. In this connection, and in connection with Secs. 4 and 5 of this paper I must recall the paper by Gross[51] on the classical Dirac-photon system, together with the quantum investigations of Refs. 52–55.

ACKNOWLEDGMENTS

The main ideas of the paper originated during the author's stay in Marseille–Luminy arranged by Dr. Ph. Combe, Dr. R. Rodriguez, Dr. M. Sirugue-Collin, and Dr. M. Sirugue, to whom I am greatly indebted. I have gained much from conversations with Professor A. Luther in Copenhagen, and last but not least with Professor A. Swieca, who indicated to me a possible need for "dyonization" in the Bose description of fermions in 1 + 3 dimensions.

[1] J-M. Levy-Leblond, Nuovo Cimento **4**, 96 (1974).
[2] S. I. Rubinov and J. B. Keller, Phys. Rev. **131**, 2783 (1963).
[3] I. Białynicki-Birula, Acta Phys. Austr. Suppl. 18 (1977).
[4] J. M. Souriau, Ann. Inst. H. Poincare **20**, 315 (1974).
[5] W. G. Dixon, Nuovo Cimento **38**, 1626 (1964).
[6] A. J. Hanson and T. Regge, Ann. Phys. (NY) **87**, 498 (1974).
[7] F. Halbwachs, *Theorie relativiste de particule a spin* (Gauthier-Villars, Paris, 1960).
[8] F. Bopp and R. Haag, Z. Naturforsch. A, 644 (1950).
[9] N. Rosen, Phys. Rev. **82**, 621 (1951).
[10] H. C. Corben, *Classical and Quantum Theories of Spinning Particles* (Holden-Day, San Francisco, 1968).
[11] J. S. Dowker and D. F. Pettengil, J. Phys. A **7**, 1527 (1974).
[12] L. Schulman, Phys. Rev. B **2**, 1558 (1968).
[13] L. Schulman, Nucl. Phys. B **18**, 595 (1970).
[14] J. P. Dahl, Kgl. Dansk. Vid. Selsk. Mat. Fys. Medd. **39**, No. 12 (1977).
[15] R. Casalbueni, Nuovo Cimento A **33**, 389 (1976).
[16] F. A. Berezin and M. S. Marinov, Ann. Phys. (NY) **104**, 336 (1977).
[17] F. Breitenloher (private communication).
[18] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[19] J. R. Klauder, Ann. Phys. (NY) **11**, 123 (1960).
[20] P. Garbaczewski, J. Math. Phys. **19**, 642 (1978).
[21] Ph. Combe, R. Rodriguez, M. Sirugue-Collin, and M. Sirugue, Comm. Math. Phys. **63**, 219 (1978).
[22] Ph. Combe, R. Rodriguez, R. Sirugue-Collin, and M. Sirugue, J. Math. Phys. **20**, 611 (1979).
[23] R. Rodriguez, Ph.D. thesis, Marseille-Luminy (1979) (unpublished).
[24] S. Naka, Progr. Theor. Phys. **59**, 2190 (1978).
[25] A. Jevicki and N. Papanicolau, Ann. Phys. (NY) **120**, 107 (1979).
[26] J. R. Klauder, Phys. Rev. D **19**, 2349 (1979).
[27] P. Garbaczewski, Phys. Lett. A **73**, 281 (1979).
[28] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[29] P. Garbaczewski, Phys. Rep. **36**, 65 (1978).
[30] P. Garbaczewski, J. Math. Phys. **21**, 2670 (1980).
[31] P. Garbaczewski, J. Math. Phys. **22**, 574 (1981).
[32] N. Nakanishi, Progr. Theor. Phys. **57**, 580 (1977).
[33] N. Nakanishi, Progr. Theor. Phys. **57**, 1025 (1977).
[34] L. K. Hadjivanov and P. T. Stoyanov, Rep. Math. Phys. **15**, 36 (1979).
[35] B. Schroer, Fortschr. Physik **26**, 441 (1978).
[36] V. N. Popov, *Path Integrals in Q.F.T. and Statistical Physics* (Atomizdat, Moscow, 1976).
[37] H. Bergknoff and H. B. Thacker, Phys. Rev. D **19**, 3666 (1979).
[38] P. Garbaczewski, Comm. Math. Phys. **43**, 131 (1975).
[39] P. Garbaczewski, "On quantum solitons and their classical relatives. II. 'Fermion-boson reciprocity' and the classical vs quantum problem for the sine-Gordon system," J. Math. Phys. **22**, 1272 (1981).
[40] E. A. Kuznetsov and A. V. Mikhailov, Teor. Mat. Fiz. **30**, 303 (1977).
[41] L. A. Takhtajan and L. D. Faddeev, Teor. Mat. Fiz. **24**, 163 (1974).
[42] K. R. Ito, Progr. Theor. Phys. **53**, 817 (1975).
[43] P. P. Kulish and L. D. Faddeev, Teor. Mat. Fiz. **4**, 153 (1970).
[44] J. Frölich, G. Morchio, and F. Strocchi, Ann. Phys. (NY) **119**, 241 (1979).
[45] N. Cabibbo and E. Ferrari, Nuovo Cimento **23**, 1147 (1962).
[46] M. Y. Han and L. C. Biedenharn, Nuovo Cimento A **2**, 544 (1971).
[47] A. Luther, Phys. Rep. **49**, 261 (1979).
[48] J. D. Bjorken and S. D.Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1964).
[49] C. L. Hammer and R. H. Tucker, J. Math. Phys. **12**, 1327 (1971).
[50] B. DeFacio and C. L. Hammer, J. Math. Phys. **17**, 267 (1976).
[51] L. Gross, Commun. Pure Appl. Math. **19**, 1 (1966).

Reduction of the super phase space for a massless Dirac particle

Bengt E. W. Nilsson
Institute of Theoretical Physics, S-412 96 Göteborg, Sweden

(Received 21 March 1980; accepted for publication 19 December 1980)

We show how to get the reduced super phase space for a classical spinning particle and how to quantize this theory. This technique introduces a Grassmannian "Hamiltonian" and a corresponding Grassmannian "time" parameter.

PACS numbers: 11.10.Ef, 11.30.Pb

I. INTRODUCTION

In 1950[1] Dirac showed how to quantize a theory described by a singular Lagrangian. The first step is to obtain the Hamiltonian formulation.[1,2] It is in this step that the constraints[1,2] appear, which tells us that the number of degrees of freedom actually are fewer than the naive phase space would indicate. There are two ways to find the quantized theory. Since only physical degrees of freedom ought to be quantized, one should eliminate the first class[2] constraints by adding gauge conditions, forming Dirac brackets, and reduce the phase space to the one of only physical degrees of freedom. We have then obtained the reduced phase space.[3] Quantization is now straightforward.

Another way is to keep the first class constraints and implement them on the quantum level as conditions on the quantum states. This latter method is the simplest one and has mostly been used. Because of its simplicity it became the one used to prove that a locally supersymmetric field theory in one dimension quantum mechanically corresponds to a Dirac particle.[4,5] In this paper we will expound on the first method for this theory and emphasize the geometrical way of understanding the elimination of first class constraints acting as restrictions on a supermanifold.[6,7] This supermanifold is in our case a manifold with a graded symplectic structure imposed on it, i.e., a super phase space (see Sec. 3).

The goal of this paper is to solve the geometrical problem of finding the reduced (physical) phase space and the equations of evolution on it and to find the quantum theory.

In Sec. 2 we will describe the method used so far in describing the Dirac particle. Our alternative way will be developed in Sec. 3 while we treat the quantization in Sec. 4. We end in Sec. 5 with a few remarks.

2. THE ACTION FOR A SPINNING PARTICLE

When we obtain the Hamiltonian formulation of a theory given by $L=(2V)^{-1}\dot{x}^{\mu}\dot{x}_{\mu}$ [5] (the Lagrangian for a spinless particle), a secondary constraint $P^2\approx 0$ [2] appears that becomes the Klein–Gordon equation after quantization. V is the "einbein" field introduced to get reparametrization invariance and $\dot{x}^{\mu}=(d/d\tau)x^{\mu}$. This Lagrangian can be generalized to describe a spinning particle by going to a superspace

$$x^{\mu}(\tau)\rightarrow y^{\mu}(\tau,\alpha)=x^{\mu}(\tau)+i\alpha\psi^{\mu}(\tau)R^{1/2}, \tag{2.1}$$

where α and ψ^{μ} are real anticommuting variables[5,8] and R an arbitrary constant of dimension length. The superspace Lagrangian[5,8,9] becomes

$$\mathscr{L}(\tau,\alpha)=-\tfrac{1}{2}i\mathrm{Ber}(h_M{}^A)(h^{-1})_A{}^M\partial_M y(h^{-1})_B{}^N\partial_N y\eta^{AB}. \tag{2.2}$$

$h_M{}^A$ is the zweibeinfield (for details see Ref. 5; note that the determinant is here changed to "Ber": Berezinian, the graded determinant). Already at this level we find the anticommuting "time" parameter. Evaluation of Ber $(h_M{}^A)$ and an introduction of $d\equiv(h^{-1})_2{}^2$, $\gamma\equiv(h^{-1})_2{}^1$, $\dot{y}\equiv dy/d\tau$, and $y'\equiv dy/d\alpha$ give

$$\mathscr{L}(\tau,\alpha)=-\tfrac{1}{2}i(\gamma d\dot{y}^2+d^2\dot{y}y'). \tag{2.3}$$

If we integrate over α and choose the gauges $\gamma'=iR^{-1/2}$ and $d'=0$, we get

$$\mathscr{L}(\tau)=\frac{1}{2V}\dot{x}^2-\tfrac{1}{2}\psi\dot{\psi}-i\frac{W}{V}\psi\cdot\dot{x}. \tag{2.4}$$

All quantities are real fields in τ. V is commuting and W anticommuting. From this Lagrangian it is straightforward to obtain the Klein–Gordon and Dirac equations using the method of imposing constraints on the quantum states. This is done in Ref. 5.

When trying to find the reduced phase space, in order to quantize only physical degrees of freedom, of this Lagrangian, two obstacles arise. First, it will turn out to be impossible to find a set of canonical variables x^i and P_i (space parts of x^{μ}, P_{μ}) for a particle with spin; see Ref. 10. Second,[9] since $\{\psi^{\mu},\psi^{\nu}\}_{\mathrm{PB}}=i\eta^{\mu\nu}$ will lead to $[\hat{\psi}^{\mu},\hat{\psi}^{\nu}]_+=2\eta^{\mu\nu}$ in the quantum case, it is impossible to define states such that $\hat{\psi}^{\mu}|\psi\rangle=\psi^{\mu}|\psi\rangle$ for all $\hat{\psi}^{\mu}$. In order to circumvent these two obstacles, we will change variables to

$$x^{11}=(1/\sqrt{2})(x^0+x^3),\quad x^{22}=(1/\sqrt{2})(x^0-x^3),$$
$$x^{12}=(1/\sqrt{2})(x^1+ix^2),\quad x^{21}=(1/\sqrt{2})(x^1-ix^2) \tag{2.5}$$

and

$$\theta^1=\tfrac{1}{2}(\psi^0+\psi^3),\quad \theta^2=\tfrac{1}{2}(\psi^0-\psi^3),$$
$$\theta^3=\tfrac{1}{2}(\psi^1+i\psi^2),\quad \theta^{3*}=\tfrac{1}{2}(\psi^1-i\psi^2). \tag{2.6}$$

The Lagrangian becomes

$$L(\tau)=(1/V)(\dot{x}^{11}\dot{x}^{22}-\dot{x}^{12}\dot{x}^{21})$$
$$-i(\theta^1\dot{\theta}^2+\theta^2\dot{\theta}^1-\theta^3\dot{\theta}^{3*}-\theta^{3*}\dot{\theta}^3)-i\sqrt{2}$$
$$\times(W/V)(\dot{\theta}^1\dot{x}^{22}+\theta^2\dot{x}^{11}-\theta^{3*}\dot{x}^{12}-\theta^3\dot{x}^{21}). \tag{2.7}$$

The naive Hamiltonian[1] resulting from (2.7) is

$$H_0=V(P_{11}P_{22}-P_{12}P_{21})$$
$$+i\sqrt{2}W(\theta^1P_{11}+\theta^2P_{22}+\theta^3P_{12}+\theta^{3*}P_{21}). \tag{2.8}$$

 0022-2488/82/030451-05$02.50 © 1982 American Institute of Physics

If we require that the primary first class constraints[1,2]

$$P_V \equiv \frac{\partial L}{\partial \dot{V}} \approx 0, \quad P_W \equiv \frac{\partial L}{\partial \dot{\omega}} \approx 0$$

be constant in time, we must also demand

$$P^2 \equiv P_{11}P_{22} - P_{12}P_{21} \approx 0 \tag{2.9}$$

and

$$\not{P} \equiv \theta^1 P_{11} + \theta^2 P_{22} + \theta^3 P_{12} + \theta^{3*} P_{21} \approx 0. \tag{2.10}$$

(2.7) gives also rise to four second class constraints, $P_i = i\theta^2$ etc., which have been eliminated by Dirac brackets. The fundamental brackets[4] are now

$$\{x^{\alpha\beta}, P_{\gamma\delta}\} = \delta^\alpha_\gamma \delta^\beta_\delta, \tag{2.11}$$

$$\{\theta^1, \theta^2\} = i/2, \tag{2.12}$$

$$\{\theta^3, \theta^{3*}\} = -i/2. \tag{2.13}$$

The rest is zero. (P_V and P_W are eliminated by the gauge choices $V \approx V_0$ and $W \approx W_0$.)

Thus, for functions f of $x^{\alpha\beta}$, $P_{\alpha\beta}$, $\theta^1, \theta^2, \theta^3$, and θ^{3*} their equations of motion are

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \{H, f\}, \tag{2.14}$$

where

$$H = V_0 P^2 + i(\sqrt{2}) W_0 \not{P}. \tag{2.15}$$

The two real constraints $P^2 \approx 0$ and $\not{P} \approx 0$ constitute a $(1+1)$-dimensional graded Lie algebra

$$\{P^2, P^2\} = 0, \quad \{P^2, \not{P}\} = 0, \quad \{\not{P}, \not{P}\} = iP^2, \tag{2.16}$$

denoted $\tilde{G}$. (2.16) gives $\tilde{G}$ realized on the phase space P defined by (2.11)–(2.13).

The final step before quantization of the reduced super phase space P_c (P_c is defined by $P^2 = 0$ and $\not{P} = 0$) will be to choose gauge conditions[10] to $P^2 \approx 0$ and $\not{P} \approx 0$ and eliminate them by forming Dirac brackets. This would, however, mean that the only equation of motion left is (2.14) and the part of P_c describing the spin of the particle will not be dynamically involved. The reason for this is that a gauge condition to $P^2 \approx 0$ will force W_0 to be zero, hence putting to zero the part of H that depends of the anticommuting quantities. To obtain the Dirac equation, this simple generalization does not seem to work. The next section shows how a more geometrical treatment can help us to reach the goal.

3. THE GEOMETRICAL VIEW

As we saw in Sec. 2 the Dirac treatment of the Lagrangian (2.2) gives rise to two constraints $P^2 \approx 0$ and $\not{P} \approx 0$ defined in (2.9) and (2.10). To get a clear picture of the elimination of these two constraints, we feel a more geometrical approach would be most fruitful.

In doing so we start with the phase space P. Consider then the action of G on this space, where G is an exponentiated group[7] having $\tilde{G}$ as its Lie algebra. P is a graded manifold spanned by coordinates

$$z^A = (x^{\alpha\beta}, P_{\alpha\beta}, \theta^1, \theta^2, \theta^3, \theta^{3*}). \tag{3.1}$$

Since the gauge group G (or rather its Lie algebra) is defined by (2.16), i.e., in terms of Poisson brackets defined on P, the action of G on P is already given: $g \in G$ is defined to be[7]

$$g(a,\alpha) \equiv e^{a\tilde{P}^2 + i\alpha\tilde{\not{P}}}, \quad \tilde{\not{P}}\tilde{\not{P}} = \tfrac{1}{2} i \tilde{P}^2, \tag{3.2}$$

where

$$\tilde{P}^2 \equiv \{P^2, \}_{PB} = J^{AB} \frac{\partial P^2}{\partial z^B} \frac{\partial}{\partial z^A} \in T_z(P) \tag{3.3}$$

and

$$\tilde{\not{P}} \equiv \{\not{P}, \}_{PB} = J^{AB} \frac{\partial \not{P}}{\partial z^B} \frac{\partial}{\partial z^A} \in T_z(P) \tag{3.4}$$

and

$$J^{AB} = \{z^B, z^A\}_{PB} = \begin{pmatrix} 0 & -1_4 & 0 \\ 1_4 & 0 & 0 \\ 0 & 0 & C_4 \end{pmatrix},$$

where

$$C_4 = \begin{pmatrix} 0 & c & 0 & 0 \\ c & 0 & 0 & 0 \\ 0 & 0 & 0 & c^* \\ 0 & 0 & c & 0 \end{pmatrix} \text{ and } c = i/2. \tag{3.5}$$

J^{AB} defines the graded symplectic structure of $T_z(P)$, the tangent space to P at the point z. A more compact notation is $\bar{a} = (a^i) = (a, \alpha)$ and $D_i = (\tilde{P}^2, i\tilde{\not{P}})$ giving $g(\bar{a}) = \exp(a^i D_i)$. From each point z_0^A in P the group G defines an orbit:

$$G \cdot z_0 \equiv \{z \in P : \exists\, g \in G : g z_0 = z\}$$

(see Fig. 1).

Furthermore, if z_0^A lies in the hypersurface P_c in P defined by $P^2 = 0$ and $\not{P} = 0$, then the whole orbit will be in P_c. This means that P and P_c will be split up into $(1+1)$-dimension superspaces parametrized by $\bar{a} = (a, \alpha)$. The equations governing these superspaces are

$$\frac{df}{da} = \frac{\partial f}{\partial a} + \{P^2, f\} \tag{3.6}$$

and

$$\left(\frac{d}{d\alpha} + \frac{i}{2}\alpha\frac{d}{da}\right) f = \left(\frac{\partial}{\partial \alpha} + \frac{i}{2}\alpha\frac{\partial}{\partial a}\right) f + \{\not{P}, f\}. \tag{3.7}$$

This second equation we get by realizing the gauge group G on its group space (a,α), or, stated differently, squaring (3.7), should give (3.6). The left-hand sides are just $\tilde{G}$ realized on the space (a,α). It is possible to prove that the orbits consist of disjoint sets of points, and thus it is possible to form a quotient space defined as the space of orbits. If one considers only points on P_c, the quotient space will be isomorphic to the "physical phase space", that is, the space of only physical

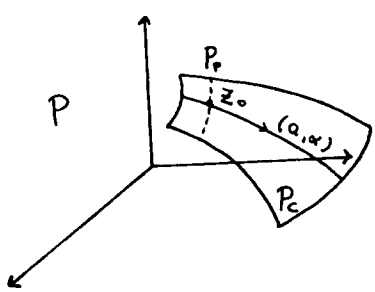

FIG. 1. The physical phase space P_p as the quotient space P_c/Gz.

degrees of freedom , the reduced phase space.

Thus the problem will be to choose the physical phase space as a subspace of P_c. This is done by selecting a set of gauge choices which will tell how the physical phase space is oriented relative to the orbits at each point. If any orbit at any point is tangent to the physical phase space, that orientation will not be a possible one. We choose the gauges ϕ_i, $i = 1,2$:

$$\phi_1 = x^{11} - a \approx 0 \tag{3.8}$$

and

$$\phi_2 = \theta^1 - \alpha \approx 0. \tag{3.9}$$

In order to preserve the orientation given by these gauge choices, (3.8) and (3.9) must not change along an orbit, i.e. (at least not if the orbit is in P_c),

$$\frac{d}{da}\phi_i \approx 0, \quad \left(\frac{d}{d\alpha} + \frac{i}{2}\alpha\frac{d}{da}\right)\phi_i \approx 0. \tag{3.10}$$

These conditions can only be satisfied if the action of the gauge group on P is modified. It can be seen that a special type of changes is allowed by pointing out that the only important concept is the geometrical (coordinate independent) $(1+1)$-dimensional superorbit not its parametrization. This view makes it possible to rescale (even locally) each Lie algebra element; i.e., it is possible to change $\tilde{P}^2$ to $F(z)\tilde{P}^2$ and $\tilde{\not{P}}$ to $F'(z)\tilde{\not{P}}$ provided it does not change the algebra [note that $\tilde{P}^2$ and $F(z)\tilde{P}^2$ point in the same direction in the tangent space of P_c at z]. In our case F and F' are essentially powers of $1/P_{22}$.

In terms of new rescaled generators, $g \in G_p$ becomes

$$g(a',\alpha') = \exp\left[a'\frac{\tilde{P}^2}{(-P_{22})} + i\alpha'\frac{\tilde{\not{P}}_{22}}{(-P_{22})}\right], \quad \bar{a}' = -P_{22}\bar{a}$$

and

$$\frac{d}{d\alpha'}f = \frac{\partial}{\partial a'}f - \frac{1}{P_{22}}\{P^2, f\}$$

and

$$\begin{aligned}
&-\frac{1}{P_{22}}\{\not{P}, f\} \\
&= -\frac{1}{P_{22}}\left(\frac{d}{d\alpha} + \frac{i}{2}\alpha\frac{d}{da}\right)f + \frac{1}{P_{22}}\left(\frac{\partial}{\partial\alpha} + \frac{i}{2}\alpha\frac{\partial}{\partial a}\right)f \\
&= \left(\frac{d}{da'} - \frac{i}{2}\frac{\alpha'}{P_{22}}\frac{d}{da'}\right)f - \left(\frac{\partial}{\partial\alpha'} - \frac{i}{2}\frac{\alpha'}{P_{22}}\frac{\partial}{\partial a'}\right)f.
\end{aligned}$$

If we drop the primes, we get

$$\frac{d}{da}f = \frac{\partial f}{\partial a} - \{P^2, f\}, \tag{3.11}$$

$$\left(\frac{d}{d\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{d}{da}\right)f = \left(\frac{\partial}{\partial\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{\partial}{\partial a}\right)f - \frac{1}{P_{22}}\{\not{P}, f\}. \tag{3.12}$$

Now

$$\frac{d}{da}\phi_i \approx 0, \quad \left(\frac{d}{d\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{d}{da}\right)\phi_i \approx 0,$$

provided the gauge choices are

$$\phi_1 = x^{11} - a \approx 0, \quad \phi_2 = \theta^1 + \tfrac{1}{2}i\alpha \approx 0. \tag{3.13}$$

The final task before quantization is to express (3.11) and (3.12) in terms of the coordinates chosen to parametrize P_p, the physical phase space. These are $z^m = (x^{22}, x^{12}, x^{21}, P_{22}, P_{12}, P_{21}, \theta^3, \theta^{3*})$. Let ϕ^i denote $(x^{11} - a, \theta^1 + i/2\alpha, P^2, \not{P})$. The cotangent space of P at z, $T_z{}^*(P)$, is spanned by dz^A These can now be rotated into the 1-forms dz^m and $d\phi^i$ by means of a z^A-dependent nonsingular matrix:

$$dz^A = dz^m E_m{}^A(z) + d\phi^i E_i{}^A(z) \equiv dz^M E_M{}^A(z),$$

where $E_M{}^A$ is a matrix which has as the upper left part [(8×8)-dimensional] a matrix belonging to OSp(2,6).

Let us express all functions $f \in C^\infty(P)$ as functions of z^m and $\phi^{i\,11}$

$$f(z^A, a, \alpha) = F(z^m, \phi^i, a, \alpha).$$

Then we get ($\partial'/\partial a$ means partial derivative of a on F)

$$\begin{aligned}
\frac{\partial f}{\partial a} &= \frac{\partial' F}{\partial a} + \frac{\partial\phi^i}{\partial a}\frac{\partial F}{\partial\phi^i} + \frac{\partial z^m}{\partial a}\frac{\partial F}{\partial z^m} \\
&= \frac{\partial' F}{\partial a} + \frac{\partial(x^{11} - a)}{\partial a}\frac{\partial F}{\partial(x^{11} - a)} + \frac{\partial z^m}{\partial a}\frac{\partial F}{\partial z^m}.
\end{aligned}$$

Thus (3.11) becomes

$$\begin{aligned}
\frac{dF}{da} &= \frac{df}{da} \\
&= \frac{\partial' F}{\partial a} + \frac{\partial F}{\partial(x^{11} - a)} + \frac{\partial z^m}{\partial a}\frac{\partial F}{\partial z^m} - \frac{1}{P_{22}}\{P^2, f\} \\
&= \frac{\partial' F}{\partial a} - \frac{\partial F}{\partial(x^{11} - a)} + \frac{\partial z^m}{\partial a}\frac{\partial F}{\partial z^m} \\
&\quad + \frac{1}{P_{22}}\frac{\partial P^2}{\partial P_{\alpha\beta}}\frac{\partial z^m}{\partial x^{\alpha\beta}}\frac{\partial F}{\partial z^m} + \frac{\partial F}{\partial(x^{11} - a)}
\end{aligned}$$

and

$$\frac{dF}{da} = \frac{\partial' F}{\partial a} + \left(\frac{\partial z^m}{\partial a} - \frac{1}{P_{22}}\{P^2, z^m\}\right)\frac{\partial F}{\partial z^m}. \tag{3.14}$$

On physical phase space P_p ($\phi^i \equiv 0$), we must have

$$\frac{dF}{da} = \frac{\partial' F}{\partial a} + J^{mn}\frac{\partial \mathscr{H}_{P^2}}{\partial z^n}\frac{\partial F}{\partial z^m}, \tag{3.15}$$

which allows us to identify (put $F = z^m$)

$$\mathscr{H}_{P^2} = P_{12}P_{21}/P_{22} \quad \text{(also denoted } \mathscr{H}_a\text{)}. \tag{3.16}$$

The same procedure for (3.12) gives

$$\begin{aligned}
&\left(\frac{d}{d\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{d}{da}\right)F \\
&= \left(\frac{\partial'}{\partial\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{\partial}{\partial a}\right)F + \left(\frac{\theta^1}{P_{22}} + \frac{i}{2}\frac{\alpha}{P_{22}}\right)\frac{\partial F}{\partial(x^{11} - a)} \\
&\quad + \left[\left(\frac{\partial}{\partial\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{\partial}{\partial a}\right)z^m - \frac{1}{P_{22}}\{\not{P}, z^m\}\right]\frac{\partial F}{\partial z^m}.
\end{aligned}$$

Restricted to P_p, this should be written as

$$\begin{aligned}
\left(\frac{d}{d\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{d}{da}\right)F &= \left(\frac{\partial'}{\partial\alpha} - \frac{i}{2}\frac{\alpha}{P_{22}}\frac{\partial}{\partial a}\right)F \\
&\quad + J^{mn}\frac{\partial \mathscr{H}_{\not{P}}}{\partial z^n}\frac{\partial F}{\partial z^m}.
\end{aligned}$$

But, surprisingly enough, this proves impossible. Fortunately, another operator $\mathscr{H}_\alpha$ can be found which is defined to be the evolution operator on P_p corresponding to the equality

$$\begin{aligned}
\frac{d}{d\alpha}F &= \frac{\partial' F}{\partial\alpha} + \left(\frac{\partial z^m}{\partial\alpha} - \frac{1}{P_{22}}\{\not{P}, z^m\}\right)\frac{\partial F}{\partial z^m} \\
&\quad - \frac{i}{2}\frac{\alpha}{(P_{22})^2}\{P^2, z^m\}\frac{\partial F}{\partial z^m} = \frac{\partial' F}{\partial\alpha} + J^{mn}\frac{\partial \mathscr{H}_\alpha}{\partial z^n}\frac{\partial F}{\partial z^m}.
\end{aligned} \tag{3.17}$$

The solution is

$$\mathscr{H}_\alpha = \frac{i}{2}\alpha\frac{P_{12}P_{21}}{(P_{22})^2} - \frac{1}{P_{22}}(\theta^3 P_{12} + \theta^{3*}P_{21}). \quad (3.18)$$

In order to get a nicer result after quantization, we here change α to $-i\alpha/2 = \alpha'$ and drop the prime. Our final result is

$$\frac{d}{d\alpha}F = \frac{\partial' F}{\partial\alpha} + J^{mn}\frac{\partial\mathscr{H}_\alpha}{\partial z^n}\frac{\partial F}{\partial z^m} \quad (3.19)$$

with

$$\mathscr{H}_\alpha = -2i\alpha\frac{P_{12}P_{21}}{(P_{22})^2} - \frac{2i}{P_{22}}(\theta^3 P_{12} + \theta^{3*}P_{21}). \quad (3.20)$$

4. QUANTIZATION ON THE REDUCED SUPER PHASE SPACE

The system of equations to be quantized is

$$\frac{dF}{da} = \frac{\partial F}{\partial a} + \{\mathscr{H}_a, F\}, \quad (4.1)$$

$$\frac{dF}{d\alpha} = \frac{\partial F}{\partial\alpha} + \{\mathscr{H}_\alpha, F\}, \quad (4.2)$$

where

$$\mathscr{H}_a = P_{12}P_{21}/P_{22} \quad (4.3)$$

and

$$\mathscr{H}_\alpha = -2i\alpha\frac{P_{21}P_{12}}{(P_{22})^2} - \frac{2i}{P_{22}}(\theta^3 P_{12} + \theta^{3*}P_{21}). \quad (4.4)$$

The PB's are defined on P_p, i.e., for $f \in C^\infty(P_p)$

$$\{f,g\} \equiv \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial p} - \frac{\partial g}{\partial x}\frac{\partial f}{\partial p}\right) - \frac{i}{2}\left(\frac{\partial f}{\partial\theta^3}\frac{\partial g}{\partial\theta^{3*}} - \frac{\partial g}{\partial\theta^3}\frac{\partial f}{\partial\theta^{3*}}\right) \quad (4.5)$$

(f,g even functions) with $x \equiv (x^{22},x^{12},x^{21})$ and $p \equiv (P_{22},P_{12},P_{21})$.

The fundamental brackets are for x, p, θ^3, and θ^{3*}

$$\{x,p\} = 1, \quad \{\theta^3,\theta^{3*}\} = -i/2 \quad \text{and the rest zero.} \quad (4.6)$$

Quantization means $\{\ ,\ \}_{PB} \to -(1/i\hbar)\,[\ ,\ ]_\pm$ ($+$ for odd, $-$ for even Grassmann variables). We obtain [$\theta \to -(1/\sqrt{2})\,a$ and $\hbar = 1$]

$$[\hat{x},\hat{p}]_- = -i \quad \text{and} \quad [a^3,a^{3+}]_+ = -1. \quad (4.7)$$

Here it should be noted that the Fock space has an indefinite metric. This is nice since it seems impossible to construct three operators generating the three Pauli matrices without using an indefinite metric in the Fock space. Calling the Fock states $|r\rangle$, $r = 1,2$ (see Appendix), we get

$$\langle r|s\rangle = (\sigma^3)_{rs}, \quad \langle r|-(a^+ + a)|s\rangle = (\sigma^1)_{rs}, \quad \langle r|i(a^+ - a)|s\rangle = (\sigma^2)_{rs}. \quad (4.8)$$

The evolution equations for a quantum state $|\psi(a,\alpha)\rangle$ become (formally!)

$$i\frac{\partial}{\partial a}|\psi\rangle = \frac{\hat{P}_{12}\hat{P}_{21}}{\hat{P}_{22}}|\psi\rangle \quad (4.9)$$

and

$$i\frac{\partial}{\partial\alpha}|\psi\rangle = \left[-2i\alpha\frac{\hat{P}_{12}\hat{P}_{21}}{(\hat{P}_{22})^2} - \frac{(\sqrt{2})i}{\hat{P}_{22}}(a^3\hat{P}_{12} + a^{3+}\hat{P}_{21})\right]|\psi\rangle, \quad (4.10)$$

where the lhs of (4.10) has been chosen in comparison with (4.9). Acting from the left with a (Hilbert space) basis vector [see (A4)]

$$\langle x,(\theta^{3*})| \equiv \langle x|\cdot\langle 0|e^{a^3\theta^{3*}}$$

(see Appendix for Fock space with negative metric) and using the completeness relation (A6) and definition (A7), (4.9) becomes

$$(\partial_{22}\partial_a - \partial_{21}\partial_{12})\psi(x,(\theta^{3*}),a,\alpha) = 0. \quad (4.11)$$

Using the coordinate transformation (2.5) and the fact that (4.11) are valid only on P_p (i.e., where $x^{11} = a$), we get

$$\Box\psi(x^\mu,(\theta^{3*}),\alpha) = 0 \quad (4.12)$$

where $x^\mu \equiv (x^0,x^1,x^2,x^3)$.

The same procedure with the Dirac-like equation (4.10) could be worked through, but in this case it is easier to work entirely with Fock states (see Appendix). Thus ($\alpha \to \alpha/\sqrt{2}$) (4.10) becomes

$$\langle x,r|i\partial_\alpha|\psi(a,\alpha)\rangle = \langle x,r|\left[-i\alpha\frac{\hat{P}_{12}\hat{P}_{21}}{(\hat{P}_{22})^2} - \frac{i}{\hat{P}_{22}}(a^3\hat{P}_{12} + a^{3+}\hat{P}_{21})\right] \times \sum_{s,t}\int dx'|x',s\rangle(\sigma^3)_{st}\langle t,x'|\psi(a,\alpha)\rangle. \quad (4.13)$$

In order to show that Eq. (4.13) is nothing but the Dirac equation $\gamma^\mu\partial_\mu = 0$ and γ^μ in the chiral representation, a few steps of calculations are needed. First we notice that on $|\psi\rangle$ the operator $\hat{P}_{12}\hat{P}_{21}/\hat{P}_{22}$ is $i\partial/\partial a$ [Eq. (4.9)]. Then we (formally) multiply both sides by $\hat{P}_{22}$ and use the x-space representation for the momentum operators. We find

$$\sum_{s,t}\langle r|(\partial_{22}\partial_\alpha + \partial_a\alpha + \partial_{12}a^3 + \partial_{21}a^{3+}) \times |s\rangle(\sigma^3)_{st}\langle t|\psi(x,a,\alpha)\rangle = 0.$$

Next we change coordinates back to x^0, x^1, x^2, x^3 and use the properties of the Fock states $\alpha|r\rangle = (-1)^{r-1}|r\rangle\alpha$ and the fact that $\Sigma_s\langle r|s\rangle(-1)^{s+1}(\sigma^3)_{st} = (\sigma^3)_{rt}$. This leads to the expression [after using (4.8)]

$$\sum_t(\sigma^3)_{rt}[(\partial_0 - \partial_3)\partial_\alpha + (\partial_0 + \partial_3)\alpha]\psi_t(x,a,\alpha) + \sum_t[i(\sigma^2)_{rt}\partial_1 + i(\sigma^1)_{rt}\partial_2]\psi_t(x,a,\alpha) = 0.$$

Finally expanding $\psi_t = \phi_t + \alpha\chi_t$ and doing the identification

$$\lambda_1 = -\chi_2, \quad \lambda_2 = \phi_1, \quad \lambda_3 = \phi_2, \quad \lambda_4 = \chi_1, \quad (4.14)$$

one finds at once $\gamma^\mu\partial_\mu\lambda = 0$.

5. REMARKS

This way of deriving the wave equations for a spinning particle has a certain freedom with respect to the grassmannian properties of the superfield $\psi(x,a,\theta^{3*},\alpha)$; it can be either a homogeneous or an inhomogeneous Grassmann field.[6] The latter case is the one obtained when one uses the second method of quantization mentioned in the Introduction. We will here briefly comment on the case of using a homogeneous superfield. The Dirac spinor then splits into two Weyl

spinors with opposite Grassmannian properties. Lorentz invariance is thus not effected, and, further, a chiral transformation must use a anticommuting parameter. Another consequence of using a homogeneous superfield in a first quantized field theory would be that the odd Weyl spinor would not give any contribution to scattering cross sections, and it appears in the theory only to keep the chiral symmetry.

The massive case could be obtained if one starts from a super phase space with more Bose and Fermi dimensions and a Lagrangian with extra local invariances to get rid of these extra dimensions (see Ref. 5).

ACKNOWLEDGMENT

I am very grateful to Dr. Lars Brink for reading the manuscript.

APPENDIX

This appendix is based on Ref. 12 (see also Ref. 9). The Fock space generated by a pair of operators (a,a^+) has as a basis $|r\rangle$ $r = 1,2$ ($|1\rangle \equiv |0\rangle$, $|2\rangle \equiv a^+|0\rangle$). The completeness relation in this space is determined by the metric:

$$\mathbf{1} = |1\rangle\langle 1| - |2\rangle\langle 2| = |r\rangle(\sigma^3)_{rs}\langle s|. \tag{A1}$$

The coherent states used in Sec. 4 are defined as

$$|(\theta^3)\rangle \equiv e^{\theta^3 a^3 +}|0\rangle \tag{A2}$$

with

$$a^3|(\theta^3)\rangle = \theta^3|(\theta^3)\rangle. \tag{A3}$$

Complex conjugation gives

$$\langle(\theta^{3*})| \equiv \langle 0|e^{a^3\theta^{3*}} \tag{A4}$$

with

$$\langle(\theta^{3*})|a^{3+} = \langle(\theta^{3*})|\theta^{3*}. \tag{A5}$$

Thus (A1) can be written ($\int d\theta\ \theta \equiv 1$)

$$\mathbf{1} = \int (d\theta^3)(d\theta^{3*})e^{-\theta^3\theta^{3*}}|(\theta^3)\rangle\langle(\theta^{3*})|. \tag{A6}$$

Super wavefunctions are defined as

$$\langle x(\theta^{3*})|\psi\rangle \equiv \psi(x^{22},x^{21},x^{12},\theta^{3*}). \tag{A7}$$

The connection between the coherent states and the Fock states is

$$\langle(\theta^{3*})|1\rangle = 1,$$
$$\langle(\theta^{3*})|2\rangle = \theta^3 \tag{A8}$$

and

$$\langle x,r|\psi(a,\alpha)\rangle \equiv \psi_r(x,a,\alpha)$$
$$= \varphi_r(x,a) + \alpha\chi_r(x,a). \tag{A9}$$

[1] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University, 1964).
[2] E. C. G. Sudarshan and N. Mukunda, *Classical Mechanics—A Modern Perspective* (Wiley, New York, 1974), Chaps. 8 and 9; A. J. Hansson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Academia Nazionale dei Lincei, Rome, 1976).
[3] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin-Cummings, New York, 1978).
[4] R. Casalbuoni, Nuovo Cimento A **33**, 389 (1976).
[5] L. Brink, P. Di Vecchia, and P. Howe, Nucl. Phys. B **118**, 76 (1977); R. Marnelius and B. E. W. Nilsson, Phys. Rev. D **20**, 839 (1979).
[6] A. Rogers, ICTP/78-79/15 (ICTP, Trieste, 1979).
[7] S. Ferrara, B. Zumino, and J. Wess, Phys. Lett. B **51**, 239 (1974).
[8] R. Marnelius and B. E. W. Nilsson, Phys. Rev. D **22**, 830 (1980).
[9] R. Marnelius and B. E. W. Nilsson, Gothenburg Preprint 79-52 (1979).
[10] C. Bachas, Phys. Lett. B **29**, 62 (1978).
[11] N. Mukunda, Gothenburg Preprint 79-20 (1979).
[12] Y. Ohnuki, *Coherent States of Fermi Operators and the Path Integral, Lecture Notes* (Nagoya Univ., 1977).

Connection between the infinite sequence of Lie–Bäcklund symmetries of the Korteweg–deVries and sine-Gordon equations

P. Kaliappan [a] and M. Lakshmanan
Department of Physics, University of Madras Autonomous Post-graduate Center, Tiruchirapalli 620 020, Tamilnadu, India

(Received 3 February 1981; accepted for publication 20 February 1981)

From the observation that the infinite sequence of Lie–Bäcklund symmetries of the potential modified Korteweg -deVries (PMK–dV) and the sine–Gordon (s–G) equations are identical, it is shown that there exists a simple connection between the Lie–Bäcklund symmetries (written in the form of evolution equations) of the Korteweg–deVries (K–dV) and s–G equations. Further, this connection is similar to the one obtained by Chodos for the conserved quantities of K–dV and s–G equations. We also point out that the result of Chodos can be realized from the equality of conserved densities of PMK–dV and s–G systems.

PACS numbers: 11.30. – j, 02.30.Jr, 03.50. – z

I. INTRODUCTION

Recently Chodos[1] showed that the infinite sequence of conserved quantities for the Korteweg–deVries (K–dV) and the sine–Gordon (s–G) systems are related by a simple connection $u = -\frac{1}{4}(v_x^2 - 2iv_{xx})$, where $u(x, t)$ and $v(x, t)$ are the solutions of the K–dV and s–G equations, respectively. He raised an open question as to whether this is merely fortuitous or whether it bespeaks a more profound connection between the two systems. Having in mind this and the fact that the conserved quantities are obtainable from the symmetries of the systems, we investigated the relations between the Lie–Bäcklund (LB) symmetry groups of these two systems. It turns out that the infinitesimal LB symmetries, written in the form of evolution equations, of the potential modified K–dV (PMK–dV) and s–G equations are identical. Now the infinitesimal LB symmetries of the PMK–dV and modified Korteweg–deVries (MK–dV) are interrelated and these in turn can be connected to the K–dV equation. Thus we finally establish a connection between the infinitesimal LB symmetries of K–dV and s–G equations in (1 + 1)-dimensions and this connection is the same as that obtained by Chodos for conserved quantities.

The plan of the paper is as follows. In Sec. II we explain briefly the concepts of LB transformation groups, infinitesimal LB symmetries and the associated evolution equations and the methods of obtaining them. In Sec. III we tersely derive a sequence of LB symmetries of the K–dV, MK–dV and PMK–dV equations and the s–G equation following the recent works[2–5] on higher order LB symmetries. In Sec. IV we establish the relations among these LB symmetries between K–dV and s–G systems and show that this connection is similar to the one discussed by Chodos from the Lax representations of the two equations. In Sec. V, as a concluding remark, we point out that the Chodos result can be obtained by realizing the fact that the PMK–dV and s–G equations have the same conserved densities.

II. LB SYMMETRY GROUPS AND INFINITESIMAL SYMMETRIES

Consider a partial differential equation of the form

$$H = H(x, t, u, u_t, u_1, u_2, u_3, \cdots, u_n) = 0, \tag{1}$$

where

$$u_i = \frac{\partial^i u}{\partial x^i}, \quad i = 1, 2, \cdots, n.$$

The most general infinitesimal operator of one parameter LB groups of transformations leaving (1) invariant may be taken as[4]

$$X(\eta) = \eta \frac{\partial}{\partial u} + D_t\eta \frac{\partial}{\partial u_t} + D_tD_x\eta \frac{\partial}{\partial u_{xt}} + D_x\eta \frac{\partial}{\partial u_1} + D_x^2\eta \frac{\partial}{\partial u_2} + D_x^3\eta \frac{\partial}{\partial u_3} + \cdots + D_x^n\eta \frac{\partial}{\partial u_n}, \tag{2}$$

where η is a function of x, t, u and partial derivatives of u with respect to x and t of any arbitrary, but finite order. D_x and D_t are the total derivative operators with respect to x and t respectively defined by

$$D_x = \frac{\partial}{\partial x} + u_1 \frac{\partial}{\partial u} + u_{1,t} \frac{\partial}{\partial u_t} + u_{2,t} \frac{\partial}{\partial u_{1,t}} + u_2 \frac{\partial}{\partial u_1} + u_3 \frac{\partial}{\partial u_2} + \cdots,$$

$$(u_{1,t} = u_{xt}, u_{2,t} = u_{xxt}, \text{ etc.}),$$

with an analogous definition for D_t. Then the invariance condition of (1) with respect to the group whose infinitesimal generators are tangent vectors (2) is given by

$$X(\eta)H|_{H=0} = 0. \tag{3}$$

The condition (3) provides an algorithm for finding η. Each independent η satisfying (3) is called a LB symmetry of equation (1). In practice, one can obtain an infinite sequence of LB symmetries (if they exist) which are polynomials in u and its derivatives with respect to the variable x by finding a recursion operator,[3] Δ, which generates infinitely many symme-

[a] On leave from N. G. M. College, Pollachi 642 001, Tamilnadu, India.

 0022-2488/82/030456-04$02.50 © 1982 American Institute of Physics

tries starting from a given symmetry. It turns out that knowledge of a few lower order symmetries enable us to construct Δ.

For a partial differential equation admitting LB symmetries $\eta(u, u_1,...)$ we define "evolution equations" [see, e. g., Ref. 6, Eq. (14)] connected with η as

$$u_t = \eta(x, t, u, u_1, u_2,...). \tag{4}$$

Evidently Eq. (4) assigns a one-parameter LB group of transformations [see, e. g., Ref. 7, Eq. (10)] which leave invariant the solutions of the associated partial differential equations. Further, any Lie–Bäcklund group admitted by the given partial differential equation is also admitted by any of (4). Interestingly, one also finds that evolution equations of form (4) describing the symmetries of different nonlinear partial differential equations can have interrelations. In fact, we find such a connection between the set of equations (4) admitted by K–dV and s–G equations, and this connection is the same as that found by Chodos in relating the infinite number of conserved quantities.[1]

III. LB SYMMETRIES FOR THE K–dV AND s–G EQUATIONS

The equations which we consider are the following:

$$w_t + w_3 + \tfrac{1}{2}w_1^3 = 0, \quad \text{(PMK–dV)}, \tag{5}$$

$$q_t + q_3 + \tfrac{3}{2}q^2 q_1 = 0, \quad \text{(MK–dV)}, \tag{6}$$

$$u_t + u_3 + \tfrac{3}{2}uu_1 = 0, \quad \text{(K–dV)}, \tag{7}$$

$$v_{xt} - \sin v = 0, \quad \text{(s–G)}. \tag{8}$$

To start with, we obtain the first three lower-order symmetries and from them we construct the recursion operator which generates the infinite sequence of symmetries for the above equations.

Now Eq. (5) admits a LB symmetry $\eta(w, w_1, w_2,\cdots,w_n)$ for a finite n if and only if for every solution $w = w(x, t)$ of Eq. (5)

$$X(\eta)[\, w_t + w_3 + \tfrac{1}{2}w_1^3\,] = 0, \tag{9a}$$

or

$$D_t\eta + D_x^3\eta + \tfrac{3}{2}w_1^2 D_x\eta = 0. \tag{9b}$$

Here $X(\eta)$, D_x, and D_t are defined as in Sec. II. By using (2) in (9), the condition (9) reduces to

$$D_x^3\eta + \tfrac{3}{2}w_1^2 D_x\eta = \eta_u(\, w_3 + \tfrac{1}{2}w_1^3) + \sum_{j=1}^{n}\eta_{u_j}D_x^j(\, w_3 + \tfrac{1}{2}w_1^3), \tag{10}$$

where

$$\eta_u = \frac{\partial}{\partial u}\eta \text{ and } \eta_{u_j} = \frac{\partial}{\partial u_j}\eta.$$

One can easily find that the following are the solutions of equation (10) for η:

$$\eta^{I} = w_1, \tag{11}$$

$$\eta^{II} = w_3 + \tfrac{1}{2}w_1^3. \tag{12}$$

η^{I} and η^{II} correspond, respectively, to x and t translation symmetries of Eq. (5). Now we assume that there exists a recursion operator Δ which generates (12) from (11).

Using this Δ we may take the next LB symmetry in the form

$$\eta^{III} = w_5 + B\,(w, w_1, w_2, w_3, w_4). \tag{13}$$

Substitution of (13) in (10) results in a polynomial form in w_5, w_6, whose coefficients should vanish since for every solution of (5) the condition (10) must hold. This leads to the following determining equations for B:

$$\frac{\partial B}{\partial u_4} = 0, \quad D_x\frac{\partial B}{\partial u_3} = \tfrac{5}{2}w_1^2.$$

Solving these equations we obtain

$$B = \tfrac{5}{2}w_1^2 w_3 + F(w, w_1, w_2), \tag{14}$$

where F is arbitrary. Following the same lines of argument as before we obtain η^{III} in a recursive manner. In this way we get

$$\eta^{III} = w_5 + \tfrac{5}{2}w_1^2 w_3 + \tfrac{5}{2}w_1 w_2^2 + \tfrac{3}{8}w_1^5. \tag{15}$$

The form of (11), (12), and (15) suggests that the recursion operator Δ is of the form

$$\Delta = D_x^2 + p(w_1) + q(w_1)D_x^{-1}(w_2). \tag{16}$$

To find p and q we note that Δ generates (15) from (12) and (12) from (11), and therefore we have

$$p + \tfrac{1}{2}qw_1 = \tfrac{1}{2}w_1^2, \tag{17}$$

$$3w_1w_2^2 + \tfrac{3}{2}w_1^2w_3 + p(\, w_3 + \tfrac{1}{2}w_1^3) + q(\, \tfrac{1}{2}w_2^2 + \tfrac{1}{8}w_1^4) = \tfrac{5}{2}w_1^2w_3 + \tfrac{5}{2}w_1w_2^2 + \tfrac{3}{8}w_1^5. \tag{18}$$

Solving (17) and (18) for p and q and inserting these in (16) we find

$$\Delta = D_x^2 + w_1^2 - w_1D_x^{-1}w_2. \tag{19}$$

The proof that this Δ is indeed the recursion operator follows from Olver's work.[6]

In exactly the same way one can obtain the first three LB symmetries and the recursion operator for the MK–dV equation as

$$\eta^{I} = q_1, \tag{20}$$

$$\eta^{II} = q_3 + \tfrac{3}{2}q^2q_1, \tag{21}$$

$$\eta^{III} = q_5 + \tfrac{5}{2}q^2q_3 + 10qq_1q_2 + \tfrac{5}{2}q_1^3 + \tfrac{15}{8}q^4q_1, \tag{22}$$

and

$$\Delta = D_x^2 + q^2 + q_1D_x^{-1}q, \tag{23}$$

and for the K–dV equation as

$$\eta^{I} = u_1, \tag{24}$$

$$\eta^{II} = u_3 + \tfrac{3}{2}uu_1, \tag{25}$$

$$\eta^{III} = u_5 + \tfrac{5}{2}uu_3 + 5u_1u_2 + \tfrac{15}{8}u^2u_1, \tag{26}$$

$$\Delta = D_x^2 + u + \tfrac{1}{2}u_1D_x^{-1}. \tag{27}$$

The s–G equation (8) will admit a LB symmetry of the form $\eta(v, v_1,...,v_n)$ if and only if for every solution $v = v(x, t)$ of Eq. (8),

$$(D_tD_x)\eta - \eta\cos v = 0. \tag{28}$$

Taking $n = 5$, we get the following determining equations for η:

$$D\eta_v = 0, \tag{29a}$$

$$\eta_{vv_5} = 0, \tag{29b}$$

$$\eta_{v_5v_1} v_{1,t} + \eta_{v_5v_2} v_{2,t} + \eta_{v_5v_3} v_{3,t} + \eta_{v_5v_4} v_{4,t} + \eta_{v_5v_5} v_{5,t} = 0, \tag{29c}$$

$$\begin{aligned}(D_x\eta_{v_1} + \eta_v)v_{1,t} + (D_x\eta_{v_2} + \eta_{v_1})v_{2,t} + (D_x\eta_{v_3} + \eta_{v_2})v_{3,t} \\ + (D_x\eta_{v_4} + \eta_{v_3})v_{4,t} + (D_x\eta_{v_5} + \eta_{v_4})v_{5,t} \\ + \eta_{v_5}v_{6,t} - \eta\cos v = 0,\end{aligned} \tag{29d}$$

where

$$v_{1,t} = v_{xt},\ v_{2,t} = v_{xxt},\ \text{etc.},$$

and

$$\eta_{v_1} = \frac{\partial}{\partial v_1}\eta,\cdots$$

$$\eta_{v_5}\eta_{v_1} = \frac{\partial^2}{\partial v_5 \partial v_1}\eta,\cdots$$

with

$$\begin{aligned} v_{1,t} &= \sin v, \\ v_{2,t} &= v_1\cos v, \\ v_{3,t} &= -v_1^2\sin v + v_2\cos v, \\ v_{4,t} &= -3v_1v_2\sin v - v_1^3\cos v + v_3\cos v, \\ v_{5,t} &= -6v_1^2v_2\cos v - 4v_1v_3\sin v + v_1^4\sin v \\ &\quad - 3v_2^2\sin v + v_4\cos v. \end{aligned}$$

Solving Eq. (29), we find the following solutions for η:

$$\eta^{\mathrm{I}} = v_1, \tag{30a}$$

$$\eta^{\mathrm{II}} = v_3 + \tfrac{1}{2}v_1, \tag{30b}$$

$$\eta^{\mathrm{III}} = v_5 + \tfrac{5}{2}v_1^2v_3 + \tfrac{5}{2}v_1v_2^2 + \tfrac{3}{8}v_1^5. \tag{30c}$$

The corresponding recursion operator is found to be

$$\Delta = D_x^2 + v_1^2 - v_1D_x^{-1}v_2. \tag{31}$$

IV. RELATIONS BETWEEN INFINITESIMAL LB SYMMETRIES

The set of equations (4) assigning the symmetry groups for different systems reads as follows:

(a) PMK–dV:

$$w_t = w_1, \tag{32a}$$

$$w_t = w_3 + \tfrac{1}{2}w_1^3, \tag{32b}$$

$$w_t = w_5 + \tfrac{5}{2}w_1^2w_3 + \tfrac{5}{2}w_1w_2^2 + \tfrac{3}{8}w_1^5. \tag{32c}$$

.....

An infinite sequence of such equations follows by considering the higher-order symmetries generated by the recursion operator $\Delta = D_x^2 + w_1^2 - w_1D_x^{-1}w_2$.

(b) MK-dV:

$$q_t = q_1, \tag{33a}$$

$$q_t = q_3 + \tfrac{3}{2}q^2q_1, \tag{33b}$$

$$q_t = q_5 + \tfrac{5}{2}q^2q_3 + 10\,qq_1q_2 + \tfrac{5}{2}q_1^3 + \tfrac{15}{8}q^4q_1, \tag{33c}$$

.....

with $\Delta = D_x^2 + q^2 + q_1D_x^{-1}q$.

(c) K-dV:

$$u_t = u_1, \tag{34a}$$

$$u_t = u_3 + \tfrac{3}{2}uu_1, \tag{34b}$$

$$u_t = u_5 + \tfrac{5}{2}uu_3 + 5u_1u_2 + \tfrac{15}{8}u^2u_1, \tag{34c}$$

.....

with $\Delta = D_x^2 + u + \frac{1}{2}u_1D_x^{-1}$.

(d) s-G:

$$v_t = v_1, \tag{35a}$$

$$v_t = v_3 + \tfrac{1}{2}v_1^3, \tag{35b}$$

$$v_t = v_5 + \tfrac{5}{2}v_1^2v_3 + \tfrac{5}{2}v_1v_2^2 + \tfrac{3}{8}v_1^5, \tag{35c}$$

.....

with $\Delta = D_x^2 + v_1^2 - v_1D_x^{-1}v_2$.

We can now observe that the systems of equations (32) and (35) are identical. On the other hand, one finds that Eqs. (34) and (35) are connected by

$$u = q^2 - 2iq_1, \tag{36}$$

where u and q are solutions of Eqs. (7) and (6). This is because Eq. (33) can be obtained from (34) with the aid of the relation (36) and the fact[6] that the recursion operators of K–dV and MK–dV are connected by $(q - iD_x)\cdot\Delta_{\mathrm{MK-dV}} = \Delta_{\mathrm{K-dV}}(q - iD_x)$. For example, Eq. (34b) becomes, after the substitution of relation (36),

$$\begin{aligned}(q^2 - 2iq_1)_t = (q^2 - 2iq_1)_{xxx} \\ + (3/2)(q^2 - 2iq_1)(q^2 - 2iq_1)_x.\end{aligned}$$

This can be written as

$$2\left(q - i\frac{\partial}{\partial x}\right)\left[q_t - q_3 - \tfrac{3}{2}q^2q_1\right] = 0.$$

It follows that $q_t = q_3 + \frac{3}{2}q^2q_1$, where q is a solution of MK–dV. The transformation (36) is the same as Miura's transformations[8] to obtain MK–dV from K–dV.

It is easy to see now that the Eqs. (33) and (32) are connected by

$$q = w_1. \tag{37}$$

The above considerations show that Eqs. (34) and (35) are connected by the relation

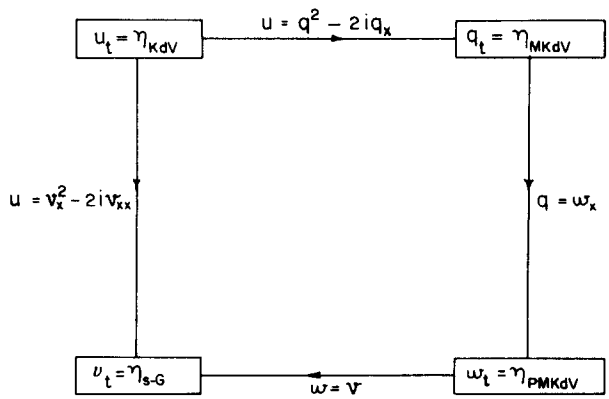

FIG. 1. Diagram showing the interrelations among the symmetries [Eqs. (32)–(35)] of the K–dV and s–G systems, written in the evolution form (4).

$$u = v_1^2 - 2iv_2, \tag{38}$$

where u and v are solutions of K–dV and s–G equations.

We now make a scale transformation $u \to -4u$ of the K–dV equation (7), which takes (7) to $u_t + u_3 - 6uu_x = 0$. Correspondingly the relation (38) becomes $u = -\frac{1}{4}(v_1^2 - 2iv_2)$, which is exactly the same as the one obtained by Chodos in relating the conserved quantities of the K–dV and s–G equations. These results are illustrated diagramatically in Fig. 1.

Further, the following relations are found to be true among the LB symmetries of the various equations:

$$\eta_{\text{K-dV}} = 2\left(q - i\frac{\partial}{\partial x}\right)\eta_{\text{MK-dV}}, \quad u = q^2 - 2iq_x, \tag{39}$$

$$\eta_{\text{MK-dV}} = \frac{\partial}{\partial x}\eta_{\text{PMK-dV}}, \quad q = w_x, \tag{40}$$

$$\eta_{\text{PMK-dV}} = \eta_{\text{s-G}}, \quad w = v. \tag{41}$$

The results lead finally to

$$\eta_{\text{K-dV}} = 2\left(v\frac{\partial}{\partial x} - i\frac{\partial^2}{\partial x^2}\right)\eta_{\text{s-G}}, \quad u = v_x^2 - 2iv_{xx}. \tag{42}$$

V. CONCLUSION

From the known forms of the conservative laws of the s–G equation $v_{xt} - \sin v = 0$ and the PMK–dV equation $w_t + w_3 + \frac{1}{2}w_1^3 = 0$ [see e. g., Ref. 9, Eqs. (2.13) and (2.14)] we can easily observe that the conserved densities of these two systems are of the same functional form. Then from the interrelations between PMK–dV→MK–dV (Miura) →K–dV, the result of Chodos is obtained. Further, the conserved quantities of the integrable equations are known to be obtainable from the LB symmetries admitted by them.[10,11] Thus we see that the connection between the conserved densities of the K–dV and s–G equations arises naturally from the connection between their LB symmetries.

ACKNOWLEDGMENT

The work of M. L. forms part of the University Grants Commission project "Differential Geometrical Approach to Nonlinear Dynamical Systems."

[1] A. Chodos, Phys. Rev. D **21**, 2818 (1980).
[2] N. H. Ibragimov and R. L. Anderson, J. Math. Anal. Appl. **59**, 145 (1977).
[3] A. S. Fokas, J. Math. Phys. **21**, 1318 (1980).
[4] G. Bluman and S. Kumei, J. Math. Phys. **21**, 1018 (1980).
[5] S. Kumei, J. Math. Phys. **16**, 2461 (1975).
[6] P. J. Olver, J. Math. Phys. **18**, 1112 (1977).
[7] N. Kh. Ibragimov and A. B. Shabat, Dokl. Akad. Nauk SSSR **244**, 57 (1979) [Sov. Phys. Dokl. **24**, 15 (1979)].
[8] R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[9] M. Wadati, H. Sanuki, and K. Konno, Prog. Theor. Phys. **53**, 419 (1975).
[10] S. Kumei, J. Math. Phys. **18**, 256 (1977).
[11] F. Magri, J. Math. Phys. **19**, 1156 (1978).

Complementary energy principles in dissipative fluids

E. W. Laedke and K. H. Spatschek
Fachbereich Physik, Universität Essen, D-4300 Essen, Federal Republic of Germany

(Received 4 August 1981; accepted for publication 7 October 1981)

Necessary and sufficient conditions for exponential stability of equilibria in dissipative fluids are discussed on the basis of energy principles. First, a known maximum principle is reformulated in a manner which is more appropriate for evaluation of actual growth rates. Secondly, complementary variational principles are presented. The latter are quite useful for qualitative estimates and numerical computations since they lead to upper and lower bounds for the exact growth rates. The results are applicable to ideal magnetohydrodynamic (MHD) cases as well as resistive plasmas and also solitary waves.

PACS numbers: 47.20. + m, 47.10. + g, 52.30. + r, 52.35.Py

I. INTRODUCTION

Energy principles are very useful in determining the stability of equilibria of dynamical systems.[1] Compared with an eigenmode analysis, they lead more easily to qualitative answers. Various formulations of energy principles have been worked out, especially in plasma physics,[2] because of their direct relevance to controlled nuclear fusion. The questions of sufficient and necessary conditions for stability have been discussed. One important contribution in this development was due to Laval *et al.*[3] who proved necessity without making use of completeness assumptions for the operators involved. And, indeed, there are many cases, for example in magnetohydrodynamics,[4] where the operators do not have a complete spectrum of eigenfunctions in the space of square integrable functions.

So far the discussion was for nondissipative systems. However, dissipative instabilities are well known and of growing interest.[5] Therefore the extension of energy principles to a large class of dissipative systems by Barston[6–9] was another important step in the development of generalized stability criteria. An extension of Barston's work to two-dimensional plasmas was done by Tasso,[10] who also discussed some generalizations to conservative systems in Lagrangian representation including dissipative terms.[11]

Usually, these energy principles are written as maximum principles for the (maximum) growth rates. Then lower bounds for the growth rates follow immediately. However, from the practical point of view it is sometimes more desirable to have upper bounds. The latter can be derived when complementary variational principles are known. Recenty,[12] we have presented such complementary energy principles in ideal magnetohydrodynamics. In this paper we generalize these findings to dissipative fluids.

The paper is organized as follows. In Sec. II, we start with some canonical form for the dynamics of the perturbations which is valid for many systems. We show that the explicit result of Barston[7] can be written in an equivalent form which is more appropriate for evaluation. However, the main reason for our reformulation is that our procedure allows us to derive complementary variational principles in a symmetric form. This is done in Sec. III, where we start with a model system in analogy to that used in Sec. II. Then it is shown that within this model system, and thereby also for the original system, the maximum growth rate follows from a minimum principle. The results are summarized in Sec. IV.

II. THE MAXIMUM PRINCIPLE

Barston[7] considered equations for small oscillations about a state of equilibrium in the form

$$N\partial_t^2\xi + D\partial_t\xi = F\xi, \tag{1}$$

for $t \geqslant 0$. Here N, D, and F are time-independent linear formally self-adjoint operators with $N \geqslant 0$ and $D \geqslant 0$. The operator D represents the dissipative forces and F the conservative forces.

It was proved that the system described by Eq. (1) is exponentially stable if and only if the system in the absence of dissipative forces is exponentially stable. Exponential instability for $\sup_\zeta \langle\zeta|F|\zeta\rangle/\langle\zeta|\zeta\rangle > 0$ is thus guaranteed under quite general conditions. Defining M as the set of functions ζ which satisfy the condition $\langle\zeta|F|\zeta\rangle > 0$ in the case of instability, the maximum growth rate γ_B can be written as[7]

$$\gamma_B = \sup_{\zeta\in M}\left[\left(\frac{1}{4}\frac{\langle\zeta|D|\zeta\rangle^2}{\langle\zeta|N|\zeta\rangle^2} + \frac{\langle\zeta|F|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}\right)^{1/2} - \frac{1}{2}\frac{\langle\zeta|D|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}\right]. \tag{2}$$

Let us briefly comment on this result. First, there is no doubt that from the theoretical point of view, the problem of stability in dissipative systems [with the assumptions underlying Eq. (1)] is thereby solved. However, for practical applications, Eq. (2) may be difficult to evaluate because of its quite complicated right-hand side. The proper choice of test functions is not obvious. Also another reason motivates us to reformulate Eq. (2). Once Eq. (2) is used as a basis for computations, we can only find lower bounds for the actual growth rate γ_B. This is obvious since numerically we approach the growth rate from below when a maximum principle is employed. Therefore the knowledge of a corresponding minimum principle is desirable. We shall derive the latter for the present problem in a symmetric way after reformulation of (2).

In order to find an equivalent form to Barston's growth rate (2), we now define the function

$$f_+(\gamma) \equiv \sup_{\zeta\in M(\gamma)} \frac{\langle\zeta|S(\gamma)|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}, \tag{3}$$

 0022-2488/82/030460-04$02.50 © 1982 American Institute of Physics

where

$$S(\gamma)\equiv F-\gamma D, \tag{4}$$

and $M(\gamma)$ is the set of functions ζ for which $\langle\zeta|S(\gamma)|\zeta\rangle>0$. Note that in the unstable case $\gamma>0$ exist so that $M(\gamma)$ is not empty. We shall prove that the greatest possible solution γ of the equation

$$\gamma^2=f_+(\gamma) \tag{5}$$

agrees with γ_B as defined in Eq. (2).

Let us first make clear the idea behind this reformulation as well as the practical advantage compared with Eq. (2). To begin with the latter statement, we recognize that $f_+(\gamma)$ is much simpler to evaluate than the right-hand side of Eq. (2), and the remaining (for example, graphical) solution of Eq. (5) is quite trivial. Our statement, that the greatest possible solution of Eq. (5) is identical to γ_B, is by no means trivial although one might expect that the model system described by

$$N\partial_t^2\xi=(F-\gamma_B D)\xi \tag{6}$$

has the same growth rate

$$\gamma^2=\sup_{\zeta\in M(\gamma_B)}\frac{\langle\zeta|F-\gamma_B D|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}, \tag{7}$$

as those systems described by Eq. (1).

To show that γ_B is a solution of Eq. (5), i.e.,

$$\gamma_B^2=f_+(\gamma_B), \tag{8}$$

we proceed as follows. First, inserting a test function $\tilde\zeta\in M$ into the argument of the supremum on the right-hand side of Eq. (2) we obtain the value $\tilde\gamma\leqslant\gamma_B$, where $\tilde\gamma$ satisfies

$$\tilde\gamma^2+\frac{\langle\tilde\zeta|D|\tilde\zeta\rangle}{\langle\tilde\zeta|N|\tilde\zeta\rangle}\tilde\gamma=\frac{\langle\tilde\zeta|F|\tilde\zeta\rangle}{\langle\tilde\zeta|N|\tilde\zeta\rangle}. \tag{9}$$

Rewriting Eq. (9), we can get a form similar to Eqs. (3) and (5), i.e.,

$$\tilde\gamma^2=\frac{\langle\tilde\zeta|S(\tilde\gamma)|\tilde\zeta\rangle}{\langle\tilde\zeta|N|\tilde\zeta\rangle}. \tag{10}$$

Thus $\tilde\zeta$ also belongs to $M(\tilde\gamma)$; since (10) is true for any test function $\tilde\zeta\in M(\tilde\gamma)$, we can write

$$\tilde\gamma^2\leqslant f_+(\tilde\gamma), \tag{11}$$

where f_+ is defined by Eq. (3). When approaching the maximum, we can define

$$\tilde\gamma=\gamma_B-\epsilon,\quad \epsilon\geqslant 0 \tag{12}$$

so that

$$(\gamma_B-\epsilon)^2\leqslant f_+(\gamma_B-\epsilon) \tag{13}$$

follows. In the limit $\epsilon\to 0$, we obtain

$$\gamma_B^2\leqslant f_+(\gamma_B). \tag{14}$$

To complete the proof of Eq. (8) we note that from Eq. (2)

$$\gamma_B+\frac{1}{2}\frac{\langle\zeta|D|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}\geqslant\left[\frac{1}{4}\frac{\langle\zeta|D|\zeta\rangle^2}{\langle\zeta|N|\zeta\rangle^2}+\frac{\langle\zeta|F|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}\right]^{1/2} \tag{15}$$

follows for any function $\zeta\in M$. Thus,

$$\gamma_B^2\geqslant\frac{\langle\zeta|S(\gamma_B)|\zeta\rangle}{\langle\zeta|N|\zeta\rangle} \tag{16}$$

or, if we maximize with respect to $\zeta\in M(\gamma_B)$,

$$\gamma_B^2\geqslant f_+(\gamma_B). \tag{17}$$

This completes the proof of Eq. (8).

To show that $\gamma\leqslant\gamma_B$ holds for any solution of Eq. (5) one proceeds as follows. First, analogous to the proof just given, one finds

$$\gamma^2=f_+(\gamma)=\sup_{\zeta\in M(\gamma)}\left[\left(\frac{1}{4}\frac{\langle\zeta|D|\zeta\rangle^2}{\langle\zeta|N|\zeta\rangle^2}+\frac{\langle\zeta|F|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}\right)^{1/2}-\frac{1}{2}\frac{\langle\zeta|D|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}\right]. \tag{18}$$

Then, because of $M(\gamma)\subset M$, $\gamma\leqslant\gamma_B$ follows immediately.

III. A MINIMUM PRINCIPLE

Let us now consider a model system with the following equation of motion for the perturbations,

$$N\partial_t^2\xi=(F-\gamma D)\xi, \tag{19}$$

where γ is a real parameter. The operators N, F, D have the same properties as those appearing in Eq. (1).

In the unstable case, perturbations governed by Eq. (19) grow with a maximum growth rate[3] γ_L,

$$\gamma_L^2=\sup_{\zeta\in M(\gamma)}\frac{\langle\zeta|S(\gamma)|\zeta\rangle}{\langle\zeta|N|\zeta\rangle}, \tag{20}$$

where $S(\gamma)=F-\gamma D$. It should be noted that for $\gamma=\gamma_B=\gamma_L$ the model system (19) has the same maximum growth rate as (1).

We now show that the same growth rate can be obtained from the minimum principle

$$\Gamma^2=\inf_{\zeta\in M(\gamma)}\frac{\langle\zeta|S(\gamma)N^{-1}S(\gamma)|\zeta\rangle}{\langle\zeta|S(\gamma)|\zeta\rangle}. \tag{21}$$

Under some extra condition [see Eq. (36) which is true if, e.g., $S(\gamma)$ has only one positive eigenvalue],

$$\Gamma^2=\gamma_L^2 \tag{22}$$

will follow.

We prove Eq. (22) in two steps. First, we show that unstable perturbations with growth rate Γ [see Eq. (21)] exist. Since, on the other hand, γ_L is the maximum growth rate,[3] $\Gamma^2\leqslant\gamma_L^2$ will follow.

Equation (19) has the following constant of motion:

$$\langle\xi|S(\gamma)N^{-1}S(\gamma)|\xi\rangle-\langle\partial_t\xi|S(\gamma)|\partial_t\xi\rangle=C. \tag{23}$$

We choose now some (initial) displacements $\xi_0\in M(\gamma)$, which satisfy

$$\partial_t\xi_0=\tilde\gamma\xi_0, \tag{24}$$

where

$$\tilde\gamma^2=\langle\xi_0|S(\gamma)N^{-1}S(\gamma)|\xi_0\rangle/\langle\xi_0|S(\gamma)|\xi_0\rangle>0. \tag{25}$$

For these perturbations $C=0$ holds.

The temporal behavior of the "kinetic energy" follows from Eqs. (19) and (23), using $C=0$,

$$\partial_t^2\langle\xi|S(\gamma)|\xi\rangle=4\langle\xi|S(\gamma)N^{-1}S(\gamma)|\xi\rangle. \tag{26}$$

From the definition of Γ [see Eq. (21)] we get for $\xi(t)\in M(\gamma)$,

$$\partial_t^2\langle\xi|S(\gamma)|\xi\rangle\geqslant 4\Gamma^2\langle\xi|S(\gamma)|\xi\rangle. \tag{27}$$

Defining T as the maximum time for which $\xi(t) \in M(\gamma)$ and $\langle \partial_t \xi | S(\gamma) | \xi \rangle > 0$ hold, we find, for $0 \leqslant t < T$, by integration of (27),

$$\partial_t \langle \xi | S(\gamma) | \xi \rangle \geqslant 2\Gamma \langle \xi | S(\gamma) | \xi \rangle. \tag{28}$$

Further integration yields, for $0 \leqslant t < T$,

$$\langle \xi | S(\gamma) | \xi \rangle \geqslant \langle \xi_0 | S(\gamma) | \xi_0 \rangle > \exp(2\Gamma t). \tag{29}$$

Thus, $T = \infty$ follows and (29) is true for any time $t \geqslant 0$. Furthermore, (29) shows that growing perturbations exist whose growth rates are at least Γ. Together with the known result that γ_L is the maximum growth rate,[3]

$$\gamma_L^2 \geqslant \Gamma^2 \tag{30}$$

is now sufficiently established.

We now demonstrate the second step of the proof of Eq. (22). We want to investigate when $\Gamma^2 \geqslant \gamma_L^2$ also holds [which together with Eq. (30) will give (22)]. For that we take an arbitrary function $f \in M(\gamma)$, such that

$$\sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | \zeta \rangle}{\langle \zeta | N | \zeta \rangle} \leqslant \sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | \zeta \rangle}{\langle \zeta | S(\gamma) | f \rangle^2} \times \sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | f \rangle^2}{\langle \zeta | N | \zeta \rangle}. \tag{31}$$

We first evaluate the second supremum on the right-hand side of Eq. (31). In doing so we note that, without loss of generality, we can normalize ζ such that

$$\langle \zeta | S(\gamma) | f \rangle^2 = \langle S(\gamma) f | N^{-1} | S(\gamma) f \rangle^2 \tag{32}$$

holds. Furthermore, defining φ through

$$\zeta = N^{-1} S(\gamma) f + \varphi, \tag{33}$$

the additional relation

$$\langle f | S(\gamma) | \varphi \rangle = 0 \tag{34}$$

follows from Eq. (32). Using Eqs. (33) and (34), we obtain

$$\sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | f \rangle^2}{\langle \zeta | N | \zeta \rangle} = \sup_{\varphi} \frac{\langle f | S(\gamma) N^{-1} S(\gamma) | f \rangle^2}{\langle f | S(\gamma) N^{-1} S(\gamma) | f \rangle + \langle \varphi | N | \varphi \rangle}. \tag{35}$$

Since N is positive definite, we get the result

$$\sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | f \rangle^2}{\langle \zeta | N | \zeta \rangle} = \langle f | S(\gamma) N^{-1} S(\gamma) | f \rangle. \tag{36}$$

Now it is obvious which additional condition we need for the first supremum on the right-hand side of Eq. (31), i.e.,

$$\sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | \zeta \rangle}{\langle \zeta | S(\gamma) | f \rangle^2} = \langle f | S(\gamma) | f \rangle^{-1}. \tag{37}$$

Anticipating for the moment Eq. (37) (see discussion below) we obtain from Eqs. (31), (36), and (37), together with Eq. (20),

$$\gamma_L^2 \leqslant \frac{\langle f | S(\gamma) N^{-1} S(\gamma) | f \rangle}{\langle f | S(\gamma) | f \rangle} \tag{38}$$

for $f \in M(\gamma)$. Since (38) is true for any $f \in M(\gamma)$, we immediately obtain

$$\gamma_L^2 \leqslant \Gamma^2. \tag{39}$$

Together with (30), then Eq. (22) follows. Thus we have derived complementary (maximum and minimum) principles for Eq. (19) under the condition (37).

The rest of the paper is devoted to a discussion of Eq. (37) and the relation of our model system (19) to the more general system (1).

First let us give a straightforward justification of Eq. (37) under the (plausible) assumption that $S(\gamma)$ has only one positive (discrete) eigenvalue. Employing steps similar to those for the evaluation of the second supremum on the right-hand side of Eq. (31), we choose [see Eq. (32)]

$$\langle \zeta | S(\gamma) | f \rangle^2 = \langle f | S(\gamma) | f \rangle^2, \tag{40}$$

because of freedom in normalization. Together with

$$\zeta = f + \varphi, \tag{41}$$

[compare Eq. (33)] again

$$\langle f | S(\gamma) | \varphi \rangle = 0 \tag{42}$$

follows. Making use of the last relation, we obtain

$$\sup_{\zeta \in M(\gamma)} \frac{\langle \zeta | S(\gamma) | \zeta \rangle}{\langle \zeta | S(\gamma) | f \rangle^2} = \sup_{\varphi} \frac{\langle f | S(\gamma) | f \rangle + \langle \varphi | S(\gamma) | \varphi \rangle}{\langle f | S(\gamma) | f \rangle^2}. \tag{43}$$

We will now prove that

$$\langle \varphi | S(\gamma) | \varphi \rangle \leqslant 0 \tag{44}$$

for functions φ satisfying Eq. (42) and $f \in M(\gamma)$, i.e.,

$$\langle f | S(\gamma) | f \rangle > 0. \tag{45}$$

We recall that for this proof we need the assumption that $S(\gamma)$ has only one discrete positive eigenvalue λ_+ with the eigenfunction e,

$$S(\gamma) e = \lambda_+ e. \tag{46}$$

We decompose the functions f and φ into parts parallel and perpendicular to e,

$$f = f_\parallel + f_\perp \tag{47}$$

and

$$\varphi = \varphi_\parallel + \varphi_\perp. \tag{48}$$

Equation (42) yields

$$\langle f_\perp | S(\gamma) | \varphi_\perp \rangle = -\lambda_+ \langle \varphi_\parallel | f_\parallel \rangle. \tag{49}$$

Making use of the Schwarz inequality, under the assumption that λ_+ is the only eigenvalue $\geqslant 0$,

$$\langle \varphi_\perp | - S(\gamma) | \varphi_\perp \rangle \langle f_\perp | - S(\gamma) | f_\perp \rangle \geqslant \langle f_\perp | - S(\gamma) | \varphi_\perp \rangle^2, \tag{50}$$

we obtain, together with Eq. (49),

$$\langle \varphi_\perp | S(\gamma) | \varphi_\perp \rangle \leqslant \lambda_+^2 \langle \varphi_\parallel | f_\parallel \rangle^2 \langle f_\perp | S(\gamma) | f_\perp \rangle^{-1}. \tag{51}$$

The right-hand side of Eq. (51) can be further estimated by making use of (45),

$$\lambda_+ \langle f_\parallel | f_\parallel \rangle + \langle f_\perp | S(\gamma) | f_\perp \rangle > 0 \tag{52}$$

or

$$\langle f_\perp | S(\gamma) | f_\perp \rangle^{-1} < -\lambda_+^{-1} \langle f_\parallel | f_\parallel \rangle^{-1}. \tag{53}$$

Inserting (53) into (51), we get

$$\langle \varphi | S(\gamma) | \varphi \rangle = \lambda_+ \langle \varphi_\parallel | \varphi_\parallel \rangle + \langle \varphi_\perp | S(\gamma) | \varphi_\perp \rangle \leqslant \lambda_+ \langle \varphi_\parallel | \varphi_\parallel \rangle - \lambda_+ \langle \varphi_\parallel | f_\parallel \rangle^2 \langle f_\parallel | f_\parallel \rangle^{-1} = 0, \tag{54}$$

i.e., the desired relation (44). Now we know that the supremum on the right-hand side of Eq. (43) is realized for $\varphi = 0$, and Eq. (43) leads to Eq. (37).

After completing the proof of

$$\gamma_L^2 = \Gamma^2, \tag{55}$$

if $S(\gamma)$ has only one positive eigenvalue, we summarize our findings up to this point. For a nondissipative system [described by Eq. (19)] with a parameter-dependent conservative force, we have shown that the maximum growth rate can be determined by either a maximum or a minimum principle. The derivation of the latter was the main aim of the first part of this section.

We now return to the original equation (1). Comparing Eq. (5) with Eq. (20) we see that for

$$\gamma = \gamma_L \tag{56}$$

the growth rate calculated from Eq. (20) agrees with the maximum growth rate for the system (1) (see Sec. II). Just now we have proved that the *same* growth rate can also be calculated from the minimum principle (21), provided $S(\Gamma)$ has only one positive eigenvalue. However, we want to emphasize that no assumptions regarding the negative eigenvalues and the completeness of the operators were made. Thus Eq. (21) is the minimum principle for Eq. (1) which we looked for.

If $S(\Gamma)$ has more than one discrete positive eigenvalue, we can use a simple transformation

$$\xi = \eta e^{\lambda t}, \tag{57}$$

where λ is a positive parameter, to rewrite Eq. (1) in the form

$$N\partial_t^2\eta + \bar{D}\partial_t\eta = \bar{F}\eta, \tag{58}$$

where

$$\bar{D} = D + 2\lambda N \tag{59}$$

and

$$\bar{F} = F - \lambda D - \lambda^2 N. \tag{60}$$

Note that $\bar{D}$ and $\bar{F}$ have the same definiteness properties as D and F, respectively. However, the eigenvalues of $\bar{S}$ are shifted to lower values, compared with the corresponding eigenvalues of S, when we again define

$$\bar{S}(\gamma) = \bar{F} - \gamma\bar{D}, \tag{61}$$

since

$$\bar{D} > 0. \tag{62}$$

Thus, assuming that the positive eigenvalues are discrete, we can always find a shift λ such that $\bar{S}(\Gamma)$ has only one positive eigenvalue.

Having in mind such a transformation when F has more than one discrete positive eigenvalue, we find Eq. (37) reasonably justified. Nevertheless, one can imagine cases (where the continuum is involved) such that (37) need not be true. These situations are not covered by the present investigation. However, we note that in many applications one has Schrödinger-type operators for which the bounded states belong to discrete eigenvalues so that our procedure is applicable.

IV. SUMMARY

In this paper, we have derived a minimum principle for dissipative fluids which supplements the known maximum principle. The latter we have also reformulated in a way which is more appropriate for numerical computations.

The validity of the minimum principle is based on a condition [Eq. (37)] which should be valid in most applications. We have shown explicitly that this condition is satisfied as long as the maximum positive eigenvalue of F is not the border of the continuum.

ACKNOWLEDGMENT

This work was supported in part by the Deutsche Forschungsgemeinschaft under SFB 162.

[1] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Oxford U. P., Oxford, 1961).
[2] R. Kulsrud, "General stability theory in plasma physics," in *Proceedings of the International School of Physics "Enrico Fermi,"* Varenna, Italy, (Academic, New York, 1964).
[3] G. Laval, C. Mercier, and R. Pellat, Nucl. Fusion **5**, 156 (1965).
[4] D. Pfirsch and H. Tasso, Nucl. Fusion **11**, 259 (1971).
[5] H. P. Furth, J. Killeen, and M. N. Rosenbluth, Phys. Fluids **6**, 459 (1963).
[6] E. M. Barston, J. Math. Phys. **8**, 523 (1967).
[7] E. M. Barston, Commun. Pure Appl. Math. **22**, 627 (1969).
[8] E. M. Barston, Phys. Fluids **12**, 2162 (1969).
[9] E. M. Barston, J. Fluid Mech. **42**, 97 (1970).
[10] H. Tasso, Plasma Phys. **17**, 1131 (1975).
[11] H. Tasso, Z. Naturforsch. A **33**, 257 (1978).
[12] E. W. Laedke and K. H. Spatschek, Phys. Lett. A **82**, 335 (1981).

Long-range order in the spin van der Waals model

M. Howard Lee
Department of Physics, University of Georgia, Athens, Georgia 30602

(Received 1 June 1981; accepted for publication 7 October 1981)

Long-range order in the spin van der Waals model is considered when the number of spins is finite and also when infinite. We show explicitly that a finite system cannot support long-range order. An infinite system at high temperatures is found to be dominated by the entropy of degenerate states of the system and, as a result, the system behaves essentially like an ideal system. In an infinite system at low temperatures, long-range order exists, fully reflecting the spin symmetry of the Hamiltonian. For the XY-like regime ($J > J_z$), m_x is finite but m_z vanishes, where m_x and m_z are *reduced* order parameters for the transverse and longitudinal directions, respectively. For the Ising-like regime ($J < J_z$), m_x vanishes but m_z is finite. The isotropic interaction ($J_z = J$) behaves as a singularity and it must be considered separately. A physical interpretation of the behavior of long-range order is offered using the geometry of spin space.

PACS numbers: 75.10. – b, 75.10.Dg, 05.70.Jk

I. INTRODUCTION

It was shown in a recent paper, referred to here as I, that in the spin-$\frac{1}{2}$ van der Waals model, there exists an ordered state characterized by nonvanishing long-range order when $N \to \infty$, where N denotes the number of spins of the system.[1] This order appears in the evaluation of the partition function in the form of a certain maximum *total-spin* value S_0, about which the integrand for the partition function is sharply peaked. Thus one can readily obtain the partition function by expanding it about this maximum in the manner of a saddle point. The resultant partition function is asymptotically exact (i.e., exact as $N \to \infty$). The condition for the existence of a stable maximum yields a transcendental equation for S_0, which in fact turns out to be the familiar mean-field equation for the spontaneous magnetization.[2] From this equation one can deduce the critical temperature T_c and also show that $S_0 = O(N)$ for $T < T_c$.

The existence of this long-range order below T_c was demonstrated in I for the XY-like and Ising-like regimes of the spin van der Waals model *separately*. It is necessary to do so since the ordering in the XY-like regime, owing to its symmetry, is different from the ordering in the Ising-like regime. In both regimes one obtains the same equation for S_0 but with different values of T_c. Above T_c the situation is very different from the ordered state. It is not necessary to distinguish whether the system is XY-like or Ising-like. In each case the order goes as $O(N^{1/2})$, so that the order per spin vanishes as $O(N^{-1/2})$ in the asymptotic or thermodynamic limit.

It was never really shown in I—only suggested—that S_0 is connected to the spontaneous magnetization of the van der Waals model. In this paper we wish to show this connection explicitly. In particular we want to study $\langle S_x \rangle$ and $\langle S_z \rangle$ for the XY-like and Ising-like regimes of the model when $T > T_c$ and $T < T_c$ and also when N is finite and infinite. We will also study the isotropic Heisenberg case ($\lambda = 0$), which owing to its unique symmetry position with respect to the XY-like and Ising-like regimes ($\lambda > 0$ and $\lambda < 0$, respectively), must be given special consideration. One cannot obtain the Heisenberg solutions from the XY-like solutions by simply making $\lambda \to 0+$ nor from the Ising-like solutions by making $\lambda \to 0-$, where λ is a symmetry parameter defined after Eq. (2).

In these studies we find a number of interesting features, which are readily demonstrable in our soluble model. For example, we will observe in the Ising-like regime of the model how symmetry must be broken to give long-range order in the system. When off-diagonal exchange dominates (i.e., XY-like), long-range order arises in a different manner. We will also observe how an ordered state emerges as $N \to \infty$.

For most nontrivial cooperative models which exhibit a second-order transition, it is not feasible—at least, not readily feasible—to calculate, e.g., $\langle S_z \rangle$ for $T < T_c$ nor to demonstrate the above-mentioned features.[3] For those few cases where it is possible, the attendant calculations and proofs are, invariably, so exceedingly involved that they tend to obscure the underlying physics. Our results for the van der Waals model shown in this paper, however, are sufficiently elementary as to help illuminate some essentials of cooperative physics.

In Sec. II we briefly describe the spin van der Waals model, reviewing mainly physics of this model already known. In Sec. III we consider a finite system to show that there is no long-range order. In Sec. IV we consider an infinite system of anisotropic spin coupling and in Sec. V, an infinite system of isotropic spin coupling. Finally, in Sec. VI we add a few concluding remarks from the vantage of our obtained results.

II. DESCRIPTION OF THE MODEL

We shall describe briefly below the spin van der Waals model. The model is more fully described in I. Consider a system of N $\frac{1}{2}$ spins situated on sites of a regular lattice. The components of the individual spin operator at site i are denoted by s_i^x, s_i^y, and s_i^z, with $1 \leqslant i \leqslant N$, and those of the total spin by S_x, S_y, and S_z, e.g.,

$$S_\alpha = \sum_{i=1}^{N} s_i^\alpha \tag{1}$$

 0022-2488/82/030464-08$02.50 © 1982 American Institute of Physics

with $\alpha = x, y,$ or z.

The model is defined by the following Hamiltonian[4] with a conveniently chosen additive constant

$$\mathcal{H} = -N^{-1}\sum_{i\neq j}^{N}[J(s_i^x s_j^x + s_i^y s_j^y) + J_z(s_i^z s_j^z)] - \tfrac{1}{4}(2J + J_z)$$
$$= -N^{-1}(J\mathbf{S}\cdot\mathbf{S} - \lambda S_z^2), \tag{2}$$

where $\mathbf{S} = (S_x, S_y, S_z)$ and $\lambda = J - J_z$ with $J, J_z \geqslant 0$.

In this system all spins are assumed to interact pairwise and with *equal* strength, independent of their relative distances. The interaction, however, is assumed to be anisotropic with respect to spin components. We observe that the Hamiltonian is diagonal in the total spin operators, which makes the model completely soluble.

Now λ can be either positive or negative. If positive, the model will be referred to as XY-like. If negative, it will be referred to as Ising-like. The partition function depends on the sign of λ for $T < T_c$, but not for $T > T_c$.[5] The value of the critical temperature T_c also depends on the sign of λ (discussed below). The isotropic Heisenberg point ($\lambda = 0$) behaves as a singularity in the line of λ, dividing the van der Waals model into two physical regimes. Thus, one cannot analytically continue across the isotropic point from the XY side into the Ising side, nor conversely. This property is reflected in the behavior of the critical temperature of the model.

For $\lambda > 0$ (i.e., $J > J_z$), the value of T_c is determined by J only. That is, for J fixed but larger than J_z, T_c remains constant, being entirely independent of the size of J_z. Similarly for $\lambda < 0$, the reverse holds. This behavior of the critical temperature is a special feature of the van der Waals model, stemming from its spin symmetry. We recall that for the *nearest-neighbor* anisotropic Heisenberg model, the critical *exponent* has been found to behave similarly.

III. FINITE SYSTEM

When N is finite, one does not expect long-range order to exist in the system (i.e., no T_c). We show explicitly in this section that $\langle S_x\rangle$ and $\langle S_z\rangle$ indeed vanish identically. For a finite system it is evidently unimportant whether the model is XY-like or Ising-like. Our proofs here are useful mainly for showing later how an ordered state can emerge in the asymptotic limit.

We will take the ensemble average in the usual way over eigenstates of $\mathcal{H}$ weighted appropriately with the degeneracy of these states. Spin ordering can, therefore, be obtained directly as the ensemble average of S_x or S_z. One can also introduce a small external field to align the spins of the system, which may be turned off later.

A. Spin ordering in the transverse direction

We define spin ordering in the transverse direction as follows:

$$\langle S_x\rangle = Z^{-1}\mathrm{Tr}\, S_x e^{-\beta\mathcal{H}}, \tag{3}$$

where the partition function Z is given by

$$Z = \mathrm{Tr}\, e^{-\beta\mathcal{H}}. \tag{4}$$

As stated above, we take the trace over eigenstates of $\mathcal{H}$, i.e.,

$$Z = \sum_{S=0}^{1/2N}\sum_{S_z=-S}^{S} g(S)\langle E(S,S_z)|e^{-\beta\mathcal{H}}|E(S,S_z)\rangle, \tag{5}$$

where g is the degeneracy factor previously given with N assumed to be an even number

$$g(S) = \binom{N}{\frac{1}{2}N - S} - \binom{N}{\frac{1}{2}N - S - 1}, \tag{6}$$

and $|E(S,S_z)\rangle$ denotes eigenstates of $\mathcal{H}$

$$\mathcal{H}|E(S,S_z)\rangle = E(S,S_z)|E(S,S_z)\rangle \tag{7a}$$

and

$$\langle E(S,S_z)|E(S',S_z')\rangle = \delta_{SS'}\delta_{S_zS_z'}. \tag{7b}$$

Hence, the partition function is reduced to

$$Z = \sum_{S,S_z} g(S)e^{-\beta E(S,S_z)}, \tag{8}$$

where

$$E(S,S_z) = -N^{-1}[JS(S+1) - \lambda S_z^2], \tag{9}$$

and it may be evaluated by carrying out the spin summations.

We can evaluate spin ordering similarly

$$\langle S_x\rangle = Z^{-1}\sum g(S)e^{-\beta E(S,S_z)}\langle E(S,S_z)|S_x|E(S,S_z)\rangle. \tag{10}$$

Using the raising and lowering operators for S_x, we obtain

$$S_x|E(S,S_z)\rangle = \tfrac{1}{2}(S - S_z)^{1/2}(S + S_z - 1)^{1/2}|E(S,S_z + 1)\rangle + \tfrac{1}{2}(S + S_z)^{1/2}(S - S_z - 1)^{1/2}|E(S,S_z - 1)\rangle. \tag{11}$$

Hence, by the orthogonality relation [see Eq. (7b)] we have $\langle S_x\rangle = 0$ as long as N is finite.

Spin ordering may also be evaluated by turning on a small external field H along the transverse direction, i.e.,

$$\mathcal{H} \to \mathcal{H} - HS_x. \tag{12}$$

Even though $[\mathcal{H}, S_x] \neq 0$, one can obtain $\langle S_x\rangle$ by the following means:

$$\langle S_x\rangle = \beta^{-1}\frac{\partial}{\partial H}\ln Z(H=0), \tag{13}$$

where $Z(H)$ is the field-dependent partition function. The noncommutativity poses no difficulty here since only those terms linear in H need be kept.[6] One can show readily that (13) leads to the same result as shown above [Eq. (10)].

B. Spin ordering in the longitudinal direction

We define spin ordering in the longitudinal direction as follows:

$$\langle S_z\rangle = Z^{-1}\mathrm{Tr}\, S_z e^{-\beta\mathcal{H}}. \tag{14}$$

It may be evaluated similarly to the ordering in the transverse direction, viz.,

$$\langle S_z\rangle = Z^{-1}\sum g(S)S_z e^{-\beta E(S,S_z)}. \tag{15}$$

But since $E(S,S_z)$ is even in S_z, we obtain $\langle S_z\rangle = 0$.

Because S_z is a diagonal operator, it is necessary to break the up–down symmetry by introducing a small up or down field,

$$\mathscr{H} \to \mathscr{H} - HS_z, \tag{16}$$

such that

$$\langle S_z \rangle = \beta^{-1} \frac{\partial}{\partial H} \ln Z(H=0). \tag{17}$$

Since $[\mathscr{H}, S_z] = 0$, the above field-dependent partition function can be more readily evaluated here than in the transverse case. Terms linear in H, however, vanish for the same reason (i.e., oddness in S_z) yielding $\langle S_z \rangle = 0$ for N finite.

We add a few remarks. Both $\langle S_x \rangle$ and $\langle S_z \rangle$ vanish for a finite system but in different ways. In the representation in which $\mathbf{S}^2$ and S_z are diagonal, $\langle S_x \rangle$ vanishes because S_x is not diagonal, i.e., the states of $\mathbf{S}^2$ and S_z are not stationary states of S_x. In effect it is like having zero eigenvalues, and symmetry cannot be broken with null eigenvalues. Now, stationary states of $\mathscr{H}$ of finite N are definite and discrete. When S_x acts on one of these stationary states it creates a new set of states by overturning spins, which are thus different from the original state. Hence, owing to orthogonality, none of these new states can singly contribute to the ensemble average. Physically, it means that the total x spin does not have a definite value. Individual spins do not align themselves but precess in the XY plane of spin space.

In this representation $\langle S_z \rangle$ vanishes because both the positive and negative eigenvalues of S_z ($S_z = -S, \ldots, +S$) contribute equally to the ensemble average. That is, all possible stationary states of $\mathscr{H}$ are included in the ensemble. The total z spin has a definite value for each of these states for a given S but with an equally probable orientation of being up or down. One can, of course, break this symmetry by introducing an external field. But as soon as the field is removed, the *finite* system restores itself to the original state of symmetry. Whether the system can restore itself depends on whether $HN \to 0$ or ∞. For a finite system, $HN \to 0$ always as $H \to 0$ (as in our results in this section). If $HN \to \infty$ as $N \to \infty$ (first), then the system can retain broken symmetry even after the field is removed (as shown in Sec. IV), and thus an ordered state can emerge asymptotically.

In the asymptotic limit, long-range order in the transverse direction can also exist (as shown in Sec. IV) if stationary states of $\mathscr{H}$ of N spins are no longer discrete and form a nondenumerably infinite or dense set of states. When S_x acts on any one of these stationary states, it will overturn a spin (i.e., $S_z \to S_z \pm 1$). But for a very large number of spins, i.e., $S_z = O(N^x)$, with $x = O(1)$ and $N \to \infty$, the overturning will have little effect on the state. Asymptotically, S_x thus acts as a quasidiagonal operator with eigenvalues $(S^2 - S_z^2)^{1/2}$ and all the stationary states can now contribute singly to the ensemble average, unlike the case when N is finite.

Also, when N is finite, the sign of λ is unimportant to stationary states of $\mathscr{H}$. But in the asymptotic limit, where the system can order itself, the sign of λ will be found essential in sorting out appropriate stationary states of the system.

IV. INFINITE SYSTEM OF ANISOTROPIC SPIN COUPLING

In this section we consider the long-range order in the system of anisotropic spin coupling ($\lambda \neq 0$) when $N \to \infty$. It is convenient to define the following reduced long-range order parameter:

$$m_\alpha = \lim_{N \to \infty} \frac{1}{\frac{1}{2}N} \langle S_\alpha \rangle, \tag{18}$$

where $\alpha = x, y,$ or z. We shall study m_α for the XY-like and Ising-like regimes. We expect all m_α's to vanish for $T > T_c$, but some m_α's to survive for $T < T_c$.

A. m_x for $T < T_c$

For an infinite system whose stationary states of $\mathscr{H}$ form a nondenumerably dense set of states (see Appendix A), we can write

$$\langle S_x \rangle = Z^{-1} \int_0^\infty dS \int_{-S}^{S} dS_z g(S) (S^2 - S_z^2)^{1/2} e^{-\beta E(S, S_z)}, \tag{19}$$

where

$$Z = \int dS\, dS_z g(S) e^{-\beta E(S, S_z)} \tag{20}$$

and, to the leading order in N,

$$g(S) = (S/N) 2^{N+2} e^{-NW(S)} \tag{21}$$

and

$$E(S, S_z) = -N^{-1}(JS^2 - \lambda S_z^2). \tag{21a}$$

Evidently, $W(S)$ is related to the entropy of the system for a given value of S.[7]

We consider $\langle S_x \rangle$ in the XY-like regime first. For $\lambda > 0$, the S_z dependence in (19) is essentially Gaussian and the main contributions to the S_z integration come from $S_z = O(N^{1/2})$. An examination shows that the phase factor $G(S) \equiv -W(S) + \beta J (S/N)^2 + O(N^{-1})$ has a sharp maximum at $S = S_0 = O(N)$, so that $\exp NG(S)$ is sharply peaked at about $S = S_0$.

To evaluate the double integral, we write $S = S_0 + Nx$ and $S_z = N_z^{1/2}$ and expand the S integrand about $S = S_0$. We then obtain, to order $O(N^{-1})$,[8]

$$\langle S_x \rangle = S_0 + AZ^{-1} \int_{-\infty}^{\infty} dx\, e^{-ax^2} x \int_{-\infty}^{\infty} dz\, e^{-bz^2}, \tag{22}$$

where a, b, and A are some constants, all positive but otherwise unimportant to our consideration. The second term of the right-hand side of (22) vanishes. Hence, for the XY-like regime below T_c

$$m_x = m + O(N^{-1}), \tag{23}$$

where $m = 2S_0/N$, which is thus of order unity and remains finite in the asymptotic limit.[9]

Next we consider $\langle S_x \rangle$ in the Ising-like regime. For $\lambda < 0$, the S_z dependence in (19) is no longer Gaussian. As may be expected, the S_z integrand, in fact, contributes maximally at $S_z = O(N)$. The S integrand is still peaked at $S = S_0 = O(N)$. Hence, now maximal S and S_z are of the same order in N. To remove the N dependence, we write $S = Nx$ and $S_z = Nz$ and express (19) in the following way:

$$m_x = \frac{\int_0^\infty dx\, e^{NG(x)} x \int_{-x}^{x} dz (x^2 - z^2)^{1/2} e^{az^2}}{\int_0^\infty dx\, e^{NG(x)} x \int_{-x}^{x} dz\, e^{az^2}}, \tag{24}$$

where $a = \lambda'\beta N, \lambda' = -\lambda > 0$.

One can show that for $a > 0$,

$$\int_{-x}^{x} dz(x^2 - z^2)^{1/2} e^{az^2} = \tfrac{1}{2}\pi x^2 e^{(1/2)ax^2}\{I_0(\tfrac{1}{2}ax^2) - I_0'(\tfrac{1}{2}ax^2)\}, \quad (25)$$

where I_0 and I_0' are, respectively, the modified Bessel function of order 0 and its first derivative with respect to its argument.[10] Now since $a \to \infty$, we can use the asymptotic form for the Bessel function,[10] i.e., $y \to \infty$,

$$I_0(y) = \frac{e^y}{(2\pi y)^{1/2}}[1 + 1/8y + O(y^{-2})]. \quad (26)$$

Hence, for $a \to \infty$,

$$\int_{-x}^{x} dx(x^2 - z^2)^{1/2} e^{az^2} = \tfrac{1}{2}(\pi)^{1/2} a^{-3/2} x^{-1} e^{ax^2}. \quad (27)$$

Substituting (27) into (24), we again expand the x integrand of (24) about $x = x_0 = S_0/N$. First, the numerator of (24) is as follows:

$$\int_0^\infty dx\, e^{NG(x)} x\{\tfrac{1}{2}\pi^{1/2} a^{-3/2} x^{-1} e^{ax^2}\}$$
$$= \tfrac{1}{2}\pi^{1/2} a^{-3/2} \int_0^\infty dx\, e^{NG_z(x)} x \cdot x^{-1}$$
$$= \tfrac{1}{2}\pi^{1/2} a^{-3/2} B x_0^{-1}, \quad (28)$$

where $G_z(x) = -W(x) + \beta J_z x^2$ and B is a constant, a common factor also in the denominator of (24). The denominator can also be evaluated. For $a \to \infty$,

$$\int_{-x}^{x} dz\, e^{az^2} = (ax)^{-1} e^{ax^2}. \quad (29)$$

Hence,

$$\int_0^\infty dx\, e^{NG(x)} x\{(ax)^{-1} e^{ax^2}\} = (ax_0)^{-1} B. \quad (30)$$

Substituting (28) and (30) into (24), we finally obtain

$$m_x = \pi^{1/2}(4\beta\lambda' N)^{-1/2} \quad (31)$$

which, therefore, vanishes in the asymptotic limit as $O(N^{-1/2})$.

B. m_z for $T < T_c$

Following Eqs. (16) and (17), we shall first break the symmetry by introducing a small external field along the longitudinal direction and consider the field-dependent partition function. We must remember that, for our infinite system, $HN \to \infty$ as $N \to \infty$ and $H \to 0$. It is thus convenient to think of H vanishing as $H = O(N^{-x})$, where $\frac{1}{2} < x < 1$ such that, e.g., $HN^{1/2} \to 0$ as $N \to \infty$ and $H \to 0$. This condition on H ensures that the external field term represents a perturbation to the internal interaction energy [see (16)].

The field-dependent partition function for an infinite system is given by

$$Z(H) = \int dS\, dS_z g(S) e^{-\beta E(S,S_z) + \beta H S_z}. \quad (32)$$

We shall evaluate $Z(H)$ in the XY-like and Ising-like regimes and obtain $\langle S_z \rangle$ from it by extracting linear terms of H.

For $\lambda > 0$ we have

$$Z(H) = 2^{N+2} N^{-1} \int_0^\infty dS\, e^{NG(S)} S \int_{-S}^{S} dS_z e^{-\lambda\beta S_z^2/N + \beta H S_z}, \quad (33)$$

where $G(S) = -W(S) + \beta J(S/N)^2 + O(N^{-1})$. As may be seen in (33), the main contributions to the S_z integration come from $S_z = O(N^{1/2})$. This fact already indicates that m_z cannot exist for the XY-like regime. Since the S_z integrand is dominated by $S_z = O(N^{1/2})$, one can immediately show that linear terms of H in (30) vanish,[11] i.e.,

$$Z(H) = Z(0) + O(H^2). \quad (34)$$

Thus, $m_z = 0$ for the XY-like regime.[12]

For $\lambda < 0$, we have

$$Z(H) = 2^{N+2} N^{-1} \int_0^\infty dS\, e^{NG(S)} S \int_{-S}^{S} dS_z e^{\lambda'\beta S_z^2/N + \beta H S_z}, \quad (35)$$

where $\lambda' \equiv -\lambda > 0$.

As in the evaluation of (24), the main contributions to the S_z integration now come from $S_z = O(N)$; and S and S_z are of the same order of magnitude in N. Hence, one can no longer expand the field-dependent term in (32). The S_z integration may, however, be related to Dawson's integral D,[13]

$$D(y) = e^{-y^2} \int_0^y dx\, e^{x^2}. \quad (36)$$

We obtain

$$Z(H) = 2^{N+2} N^{-1} p^{-1/2} e^{-pQ^2} \int_0^\infty dS\, e^{NG(S)} S \times \{e^{p(S+Q)^2} D(p(S+Q)) + e^{p(S-Q)^2} D(p(S-Q))\}, \quad (37)$$

where $p = \lambda'\beta/N$ and $Q = HN/2\lambda'$. Since $S = O(N)$ we can use in (37) the asymptotic form for Dawson's integral $D(y)$, i.e., for $y \to \infty$,[13]

$$D(y) = \tfrac{1}{2}y + O(y^{-3}). \quad (38)$$

Thus for $N \to \infty$,

$$Z(H) = 2^{N+1} N^{-1} p^{-3/2} e^{-pQ^2} \int_0^\infty dS\, e^{NG_z(S)} S \times \left\{\frac{e^{2pSQ + pQ^2}}{S+Q} + \frac{e^{-2pSQ + pQ^2}}{S-Q}\right\}, \quad (39)$$

where $G_z(S) = -W(S) + \beta J_z (S/N)^2 + O(N^{-1})$. Now since $2pSQ = \beta HS$ and $pQ^2 = H^2 N/4\lambda'$, we can neglect pQ^2 in the argument of each of the above exponential functions inside the brace of (39). Hence for $N \to \infty$,

$$Z(H) = 2^{N+1} N^{-1} p^{-3/2} \int_0^\infty dS\, e^{NG_z(S)} S \times \left\{\frac{e^{\beta HS}}{S + HN/2\lambda'} + \frac{e^{-\beta HS}}{S - HN/2\lambda'}\right\}. \quad (40)$$

Now we can proceed to evaluate the S integration. It has been shown previously that $\exp NG_z(S)$ is sharply peaked about $S = S_0 = O(N)$. Hence, the integration may be carried out by expanding the integrand about $S = S_0$. Thus, we get the leading order as shown below:

$$Z(H) = C\left\{\frac{e^{\beta HS_0}}{S_0 + HN/2\lambda'} + \frac{e^{-\beta HS_0}}{S_0 - HN/2\lambda'}\right\}, \quad (41)$$

where C is a constant, independent of H. Clearly, the second

exponential term may be dropped (i.e., symmetry broken) since $HN \to \infty$.[14] Also, in the denominator of the first term in the right-hand side of (41), we have $S_0 + HN/2\lambda' \approx S_0$ for $N \to \infty$. Thus, we can now obtain $\langle S_z \rangle$ via (17)

$$\langle S_z \rangle = S_0, \tag{42a}$$

or

$$m_z = m \tag{42b}$$

in the Ising-like regime.

C. m_x and m_z for $T > T_c$

For $T > T_c$, we expect m_x and m_z to vanish in both the XY-like and Ising-like regimes. We shall show below that in all cases the S and S_z integrands make maximum contributions at $S = O(N^{1/2})$ and $S_z = O(N^{1/2})$ suggesting that m_x and m_z cannot exist in the asymptotic limit.

We shall consider m_x first. Then, using (19) we can express $\langle S_x \rangle$ for $T > T_c$ in the following form:

$$\langle S_x \rangle = Z^{-1} 2^{N+2} N^{-1} \int_0^\infty dS\, e^{-cS^2/N} S \times \int_{-S}^{S} dS_z (S^2 - S_z^2)^{1/2} e^{-\lambda \beta S_z^2/N}, \tag{43}$$

where $c = (2 - \beta J) > 0$.

Now whether λ is positive or not, S_z in (43) is bounded by $(-S, S)$ and the main contributions to the S integration come from $S = O(N^{1/2})$ only. Hence, clearly $\langle S_x \rangle = O(N^{1/2})$ and $m_x = 0$ in the XY-like and Ising-like regimes.

We shall next consider m_z. Then, using (32), we can express $Z(H)$ for $T > T_c$ in the following form:

$$Z(H) = 2^{N+2} N^{-1} \int_0^\infty dS\, e^{-cS^2/N} S \times \int_{-S}^{S} dS_z e^{\lambda \beta S_z^2/N + HS_z}. \tag{44}$$

As before, it is sufficient to find linear terms of H in $Z(H)$ to obtain $\langle S_z \rangle$. Again whether XY-like or Ising-like, S_z in (44) is also bounded by $(-S, S)$ and the main contributions to the S integration come from $S = O(N^{1/2})$. Hence, the field term containing HS_z may be expanded; its linear term of H, however, vanishes. Therefore $m_z = 0$ in the XY-like and Ising-like regimes.

V. INFINITE SYSTEM OF ISOTROPIC SPIN INTERACTION

Viewed through the geometry of spin space, the XY-like system may be thought of as representing an oblate configuration and the Ising-like system a prolate one. Then $\mathscr{H}$ ($\lambda = 0$) or the isotropic system represents a sphere, which is thus a unique limiting boundary for each of the nonspherical configurations. One is tempted to think that the partition function and long-range order for the isotropic system can be obtained from those of the anisotropic systems by making $\lambda = 0$. For $T > T_c$, one can indeed do so; but for $T < T_c$, one cannot. Our low-temperature solutions, as may be noted, do not permit $\lambda = 0$. In obtaining these analytic solutions we have, in fact, made use of the nonsphericity of the anisotropic systems explicitly.

We shall first obtain the partition function for $\lambda = 0$. To obtain long-range order, it is necessary to introduce an external field along some direction to align the spins. Otherwise, $\langle S_x \rangle = \langle S_y \rangle = \langle S_z \rangle = 0$ owing to its symmetry. We shall consider $\langle S_x \rangle$ and $\langle S_z \rangle$ while the spins are aligned along the longitudinal direction.

A. Partition function

The partition function for $\lambda = 0$ may be written as

$$Z = 2^{N+2} N^{-1} \int_0^\infty dS\, e^{NG(S)} S \int_{-S}^{S} dS_z. \tag{45}$$

For $T > T_c$, the S dependence is essentially Gaussian. Thus, by integrating directly we obtain

$$Z = 2^{N+1} (\pi N)^{1/2} / (2 - \beta J)^{3/2}. \tag{46}$$

One can obtain this result from the oblate or prolate partition function by making $\lambda \to 0 \pm$ (i.e., $J_z \to J$).

For $T < T_c$, the S dependence is not Gaussian. The S integrand is peaked at $S = S_0 = O(N)$. By expanding about S_0, we obtain

$$Z = 2^{N+2} N^{-1} S_0 \exp\{-NW(S_0) + \beta J N m^2/4\}. \tag{47}$$

One cannot get this result from the XY-like or Ising-like partition function by making $\lambda \to 0 \pm$. However, the free energy per spin,

$$\ln Z/N = \ln 2 - (W - \tfrac{1}{4}\beta J m^2), \tag{48}$$

can be obtained from the nonspherical free energies by taking $\lambda \to 0 \pm$.

B. m_z

We now introduce a small external field along the longitudinal direction and obtain m_z via the field-dependent partition function

$$Z(H) = 2^{N+2} N^{-1} \int_0^\infty dS\, e^{NG(S)} S \int_{-S}^{S} dS_z e^{\beta H S_z}. \tag{49}$$

Now for $T > T_c$, $S = O(N^{1/2})$. Hence there is no linear term in H. That is, $m_z = 0$ for $T > T_c$.

For $T < T_c$ we obtain

$$Z(H) = Z(0)(S_0/\beta H)\{e^{\beta H S_0} - e^{-\beta H S_0}\}, \tag{50}$$

where $Z(0) = Z$, which is field-independent and given by (47). By dropping the second exponential term in (50) since $HS_0 \to \infty$ (i.e., symmetry-breaking) and then by taking a derivative with respect to βH, we obtain

$$m_z = m + O(N^{-1}). \tag{51}$$

C. m_x

We now calculate long-range order in the transverse direction while the spins are aligned by an external field along the longitudinal direction. The field will be, as before, turned off after $N \to \infty$. Then we have

$$\langle S_x \rangle = \frac{\int dS dS_z g(S)(S^2 - S_z^2) e^{-\beta E'(S, S_z)}}{\int dS dS_z g(S) e^{-\beta E'(S, S_z)}}, \tag{52}$$

where $E'(S,S_z) = -JS^2/N - HS_z$.

For $T > T_c$, $S = S_z = O(N^{1/2})$. Hence, it follows directly from (52) that $\langle S_x \rangle = O(N^{1/2})$ and $m_x \to 0$ as $N \to \infty$.

For $T < T_c$, (52) may be written as

$$m_x = \frac{\int_0^\infty dx\, e^{NG(x)} x \int_{-x}^{x} dx (x^2 - z^2)^{1/2} e^{\gamma z}}{\int_0^\infty dx\, e^{NG(x)} x \int_{-x}^{x} dz\, e^{\gamma z}}, \tag{53}$$

where $\gamma = \beta HN$. The above is very similar to (24), differing only in the second exponential term. One can show that

$$\int_0^x dz (x^2 - z^2)^{1/2} e^{\gamma z} = \pi x^2 \{I_0(\gamma x) - I_0''(\gamma x)\}, \tag{54}$$

where I_0'' is the second derivative of the Bessel function with respect to its argument. Now since $\gamma \to \infty$, we can again use the asymptotic form for the Bessel function

$$I_0(\gamma x) - I_0''(\gamma x) = (2\pi)^{-1/2} (\gamma x)^{-3/2} e^{\gamma x}. \tag{55}$$

Hence,

$$\int_0^\infty dx\, e^{NG(x)} x \int_{-x}^{x} dz (x^2 - z^2)^{1/2} e^{\gamma z} = C (\pi/2\gamma^3)^{1/2} x_0^{-1} e^{\gamma x_0}, \tag{56}$$

where C is a constant, a common factor also in the denominator of (53). The denominator is essentially the field-dependent partition function

$$\int_0^\infty dx\, e^{NG(x)} x \int_{-x}^{x} dz\, e^{\gamma z} = C\gamma^{-1} e^{\gamma x_0}, \tag{57}$$

where we have again dropped $\exp(-\gamma x_0)$ in (57) since $\gamma \to \infty$. Thus, we obtain finally

$$m_x = (\pi x_0 / 2\beta HN)^{1/2} \to 0, \tag{58}$$

since $HN \to \infty$.

VI. CONCLUDING REMARKS

We have demonstrated in this paper that long-range order is built up with individual spins which all become aligned along some direction. For a finite number of spins, the system cannot support broken symmetry. We have shown that as soon as an external orienting field is removed, the system restores itself to the original state of symmetry. For an infinite number of spins the collective behavior of the system is vastly different. There can be an ordered state here, which is fully dependent on the symmetries of the Hamiltonian.

For high temperatures (i.e., $T > T_c$), an infinite system is dominated by the entropy of degenerate states. Thus, the interaction or even the symmetries of the Hamiltonian are unimportant. In fact, the specific heat vanishes entirely as in an ideal system.[2] This stems from the fact that if $S = O(N^{1/2})$ only, then the energy per spin vanishes asymptotically. Hence, it is as if there are no interactions at all. For this reason, the high-temperature partition function behaves smoothly as a function of λ for all values of λ, including $\lambda = 0$.

For low temperatures (i.e., $T < T_c$), we have shown that long-range order exists and it depends on λ like a step function about $\lambda = 0$ (or about the isotropy point $J_z = J$), which acts, therefore, like a singularity. We note that this behavior is merely a reflection of the spin symmetry of the Hamiltonian. Hence, if $J > J_z$, $m_x = m$ but $m_z = 0$. If $J < J_z$, $m_x = 0$ but $m_z = m$. If $J = J_z$, m_x or m_z is finite, depending on the direction of the orienting field and the other vanishes.

For long-range order to exist, it is of course necessary to break some symmetry of the system. As is customary, we do so by orienting the spins along some direction with an external field. The controlling factor turns out to be the behavior of HN (i.e., whether $HN \to 0$ or ∞ as $H \to 0$). We have insisted *a priori* on the condition that, to achieve symmetry break, $HN \to \infty$ as $N \to \infty$ and $H \to 0$. This condition has played a crucial role throughout in determining whether the quantity HS is vanishingly small or indefinitely large. In the former, symmetry break is prohibited and in the latter, symmetry break is permitted. For this condition to be applicable throughout, it was necessary to demand that $H = O(N^{-x})$ with $\frac{1}{2} < x < 1$. We have offered no justification for it other than a physical argument. Namely, if HS_x or HS_z is to be a perturbation to the Hamiltonian, the strength of H cannot be greater than what we have demanded.

In the evaluation of $\langle S_x \rangle$, the Bessel function I_0 appeared. The appearance of this function is intriguing, inviting an explanation. The origin seems to be traceable to cylindrical symmetry present in spin space. If the system is Ising-like (i.e., prolate-shaped in spin space), there is nonzero m_z—a physical axis—about which complete rotational symmetry exists. Although m_x vanishes asymptotically, its direction in spin space is arbitrary as long as it is orthogonal to the direction of m_z. We have in fact proved in I that the time-correlation function $\langle S_x S_x(t) \rangle$ has Larmor precession about the direction of m_z, which suffers critical slowdown since $m_z \to 0$ as $T \to T_c$. This situation persists up to and including the isotropic point, provided that the orienting field remains fixed along the z direction in spin space.

If the system is XY-like (i.e., oblate-shaped in spin space), the system is unable to enjoy this kind of rotational symmetry since $m_z = 0$. That is, $\langle S_x \rangle$ cannot precess, as there are no physical axes perpendicular to it. Hence, its motion resembles a succession of tumblings in spin space. We have also proved[1] that $\langle S_x S_x(t) \rangle$ in the XY-like regime does not precess and behaves smoothly across T_c. But as soon as one introduces an external field strong enough to align the spins along the z direction, then the Bessel function will be realized.

Finally, the van der Waals model is a high density limit of the nearest-neighbor anisotropic Heisenberg model (i.e., $q \to \infty$, where q is the coordination number).[4] In the limiting process the full spin symmetry of the Heisenberg model must evidently become simplified and reduced to the elementary symmetry exhibited by the van der Waals model. Thus, the physical and mathematical results of the van der Waals model are likely to be no more than a shadow of those of the Heisenberg model. Therefore, one cannot expect to learn much about the Heisenberg model from the van der Waals model. But our understanding of the simpler model can

serve as a basis for gaining insight into the more complicated model.

ACKNOWLEDGMENTS

The author is indebted to Dr. R. Dekeyser of Katholieke Universiteit Leuven, Belgium, for general ideas about the van der Waals model. It is a pleasure to thank Dr. R. L. Anderson of the University of Georgia and Dr. Doochul Kim of Seoul National University for helpful discussions on this work. A portion of this work was accomplished while the author was a visitor at Seoul National University. He wishes to thank Dr. Y. S. Koh and Dr. H. S. Song, Dean of the College of Natural Sciences and Chairman of the Department of Physics, respectively, for their hospitality during his stay. This work was supported in part by the U. S. Department of Energy under Contract No. DE-AS09-77-ER01023 and also by the SNU-AID Graduate Basic Sciences Program.

APPENDIX A: ASYMPTOTIC STATIONARY STATES

For a finite system, $\langle S_x \rangle = 0$. If $\langle S_x \rangle \neq 0$ as $N \to \infty$, symmetry break exists or, equivalently, stationary states of $\mathscr{H}$ must also become stationary states of S_x, where $[\mathscr{H}, S_x] \neq 0$. We shall term a state which becomes asymptotically stationary an asymptotic stationary state. We show below that if macroscopic eigenvalues (defined below) merge and approach a limiting value smoothly as $N \to \infty$, then symmetry break can be achieved.

If, as $N \to \infty$, S and $S_z = O(N^y)$, where $0 < y \leqslant 1$, we shall term them *macroscopic* and their eigenvalues macroscopic eigenvalues. In this Appendix we limit our consideration to macroscopic quantities only. Consider (21a) as energy per spin $\epsilon_N(S,S_z)$,

$$\epsilon_N(S,S_z) = -\{J(S/N)^2 - \lambda (S_z/N)^2\}. \tag{A1}$$

As $N \to \infty$, ϵ_N smoothly approaches a limiting value and thus its eigenstate $|\epsilon_N>$ also approaches the corresponding eigenstate smoothly.[15] Then, for fixed N and S, we define the energy difference between two adjacent levels as

$$\Delta\epsilon_N(S,S_z) = \epsilon_N(S,S_z) - \epsilon_N(S,S_z - 1). \tag{A2}$$

If N is finite, $\Delta\epsilon_N$ is finite also, indicating a discreteness of the energy levels. But as $N \to \infty$, $\Delta\epsilon_N$ being $O(N^{-1})$ becomes arbitrarily small, i.e., $\Delta\epsilon_N \to d\epsilon_N$. Hence, asymptotically, the energy behaves classical.

For a finite system whose energy levels are discrete,

$$S^{\pm}|\epsilon_N(S,S_z)\rangle = c^{\pm}(S,S_z)|\epsilon_N(S,S_z \pm 1)\rangle, \tag{A3}$$

where

$$c^{\pm}(S,S_z) = \tfrac{1}{2}(S \mp S_z)^{1/2}(S \pm S_z - 1)^{1/2} \tag{A4}$$

and $S^{\pm} = S_x \pm iS_y$. As the energy levels merge, one can Taylor expand a state, say, of (S,S'_z) about a nearby state of (S,S_z), where $(1/N)(S'_z - S_z) = O(N^{-1})$, i.e.,

$$|\epsilon_N(S,S'_z)\rangle = |\epsilon_N(S,S_z)\rangle + (S'_z - S_z)\frac{\partial}{\partial S_z}|\epsilon_N(S,S_z)\rangle + \cdots. \tag{A5}$$

Now the second expansion term may be written as

$$\frac{\partial}{\partial S_z}|\epsilon_N(S,S_z)\rangle = \frac{\partial N}{\partial S_z}\frac{\partial}{\partial N}|\epsilon_N(S,S_z)\rangle. \tag{A6}$$

By our requirement on the eigenstates, the right-hand side is well behaved, in fact, vanishing as $O(N^{-1})$. Hence,

$$\lim_{N\to\infty}|\epsilon_N(S,S'_z) = |\epsilon_N(S,S_z)\rangle + O(N^{-1}). \tag{A7}$$

It follows directly

$$\lim_{N\to\infty}\langle\epsilon_N(S,S'_z)|S_x|\epsilon_N(S,S_z)\rangle = (S^2 - S_z^2)^{1/2}. \tag{A8}$$

Our argument for $\langle S_x \rangle \neq 0$ is reminiscent of an argument due to Bogoliubov[16] for a system of weakly interacting many bosons lying in the condensed state (i.e., the zero-momentum state). He argued that $<\Psi_0(N_0)|a_0^{\pm}|\Psi_0(N_0)> \neq 0$, where $a_0^{\pm}$ are boson creation/destruction operators in the zero-momentum state, since $\Psi_0(N_0) - \Psi(N_0 \pm 1) = O(N^{-1})$, where Ψ_0 and N_0 refer, respectively, to the wavefunction and boson number in the zero-momentum state. Thus, the Bose condensed state is also a state of broken symmetry.

One can express our energy eigenstates in the coordinate representation. We note that our Hamiltonian (2) is in the form of a one-dimensional harmonic oscillator. Thus the wavefunctions will be of Hermite polynomials. Since the eigenvalues are macroscopically large, the corresponding Hermite polynomials are thus of degree $n \to \infty$ (i.e., the zeros of Hermite polynomials merging to form a continuum and the wavefunctions behaving classical as already noted). Hence, (A5) can be given an equivalent expansion in the coordinate representation.

[1] R. Dekeyser and M. H. Lee, Phys. Rev. B **19**, 265 (1979). This reference will be referred to as I.

[2] See, for example, R. Brout, *Phase Transitions* (Benjamin, New York, 1965).

[3] See, for example, E. Lieb, T. Schultz, and D. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961).

[4] C. Kittel and H. Shore, Phys. Rev. **138**, A1165 (1965); T. Niemeijer, Physica (Utrecht) **48**, 467 (1970); G. Vertogen and A. S. DeVries, *ibid.* **59**, 634 (1972); H. Falk, Phys. Rev. B **5**, 3638 (1972); D. K. Dacol, Physica (Utrecht) **100A**, 496 (1980); I. M. Kim and M. H. Lee, Phys. Rev. B **24**, 3961 (1981).

[5] The free energy per spin for an infinite system, however, does not depend on the sign of λ for all temperatures.

[6] We are taking advantage of the cyclic permutation property under the sign of trace. One cannot, however, obtain $\langle S_x^2 \rangle$ from (17) by taking a second derivative of $Z(H)$.

[7] As in finite systems, one can also obtain $\langle S_x \rangle$ by turning on a small external field in the transverse direction [see Eqs. (12) and (13)]. Since it leads to the same result obtained here, we shall not follow this approach.

[8] The two exponential functions of S and S_z in (21) are peaked at different regions of N with little overlap. We have therefore extended the limits of integration as indicated in (22). The linear term of x in (22) comes from $(S^2 - S_z^2)^{1/2}$, noting that $S_z = O(N^{1/2})$.

[9] Also, the difference in N dependence between S and S_z in (19) allows us to regard the S_z integrand as sharply peaked at about $S_z = 0$, i.e., like a delta function $\delta(S_z)$. Thus, (19) may be written as $\langle S_x \rangle = \int dS\, g(S)\, e^{-\beta E(S,S_z)} S / \int dS\, g(S)\, e^{-\beta E(S,S_z)}$ introducing little error. Now an expansion of the S integrand about $S = S_0$ leads directly to (23).

[10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Appl. Math. Ser. No. 55 (U.S. GPO, Washington, D.C., 1970), pp. 376–7.

[11] If $S_z = O(N^{1/2})$, then $HS_z \to 0$. Hence, one can expand the S_z integrand of

(33) in powers of HS_z and evaluate the resulting linear term.

[12]It has been previously shown that $\exp NG(S)$ in Eq. (33) is sharply peaked about $S = S_0 = O(N)$. Thus, the S and S_z integrations of (33) are peaked at different regions of N and one can extend the limits of the S_z integration from $(-S,S)$ to $(-\infty,\infty)$ with impunity. Then, Eq. (33) becomes a product of two independent integrals. Hence, it follows that there are no linear terms of H. If the S and S_z integrations were peaked at the same region of N, one could not of course reduce Eq. (33) to a product of two. This proves to be the case for $\langle S_z \rangle$ in the Ising-like regime.

[13]*Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. II, p. 147.

[14]The field-dependent partition function may be written as $Z(H) = 2^{N+2}N^{-1}\int dS\, e^{NG(S)}SF(S,H)$, where $F(S,H)$ is the result of carrying out the S_z integration in (35). Symmetry break takes place in $F(S,H) = F_+(S,H) + F_-(S,H)$. For $H > 0$ and $S > 0$, $F(S,H) \to F_+(S,H)$ only. Note that the above form is different from the XY-like equation (30), where the double integral reduces to a product of two independent integrals. See, also, H. E. Stanley, *Introduction to phase transitions and Critical Phenomena* (Clarendon, Oxford, 1971), p.93.

[15]In the continuous portion of the spectrum, we assume that the eigenvector is a continuous function of the eigenvalue. See, for example, E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961), p. 321.

[16]D. Pines, *The Many Body Problem* (Benjamin, Reading, MA, 1962), p. 70; N. Bogoliubov, J. Phys. (USSR) **11**, 23 (1947).